La spectroscopie proche infrarouge et le phénotypage dans les semences

22èmes Rencontres HELIOSPIR

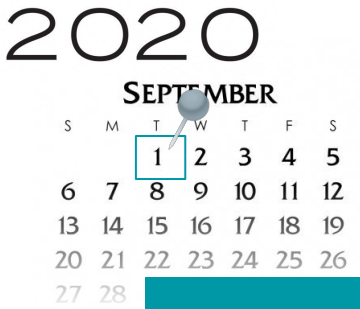
Montpellier - 24 - 25 nov. 2021

Magali Roussel - Milagros Garcia

Discovery & Technology Research

Lidea

FRESH IDEAS FOR AGRICULTURE



EURALIS
Creating seeds and trust

caussadesemences

euralis
NOURRIR VOTRE CONFIANCE

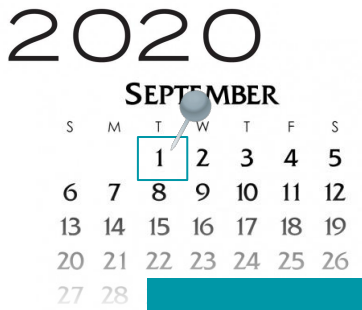
lidea

Semencier multi-espèces

- 2 000 collaborateurs
- CA = 350 millions (E)

- 17 stations de recherche - Europe
- > 30 millions d'euros investit en R&D

- 8 sites de production Europe - Ukraine - Russie
- 45 000 ha de production répartis sur 5 pays



EURALIS
Creating seeds and trust

caussadesemences

euralis
NOURRIR VOTRE CONFIANCE

Lidea

Semencier multi-espèces

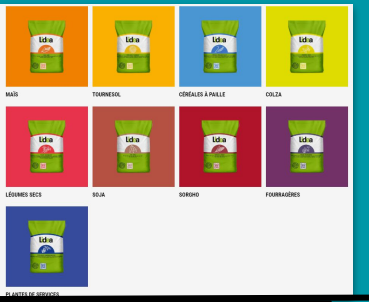
- 2 000 collaborateurs
- CA = 350 millions (E)

- 17 stations de recherche - Europe
- > 30 millions d'euros investit en R&D

- 8 sites de production Europe - Ukraine - Russie
- 45 000 ha de production répartis sur 5 pays

R&D

PRODUCTION



COMMERCIALISATION

R&D

Sélection variétale → chaque année, screening de 100aines de milliers de plantes couvrant différents environnements avec des outils de génotypage et de phénotypage

- qualité des données : homogènes / fiables / précises
- rapidité/efficacité de mise à disposition des données
- coût maîtrisé (logistique et temps de travail)

Technique de phénotypage de prédilection pour accéder à la composition biochimique = NIRS (labo + embarqué sur les machines de récolte)

En chiffres

> 15 ans d'expérience

~ 20 appareils

En chiffres

> 15 ans d'expérience

~ 20 appareils

Critères prédits- routine

LABO

> 50

Cal. commerciales / projets coll. / dvpt interne

Humidité, protéine, huile, ...

EMBARQUE

< 10

Dvpt interne + cal. commerciales

Humidité, protéine, huile, ...

En chiffres

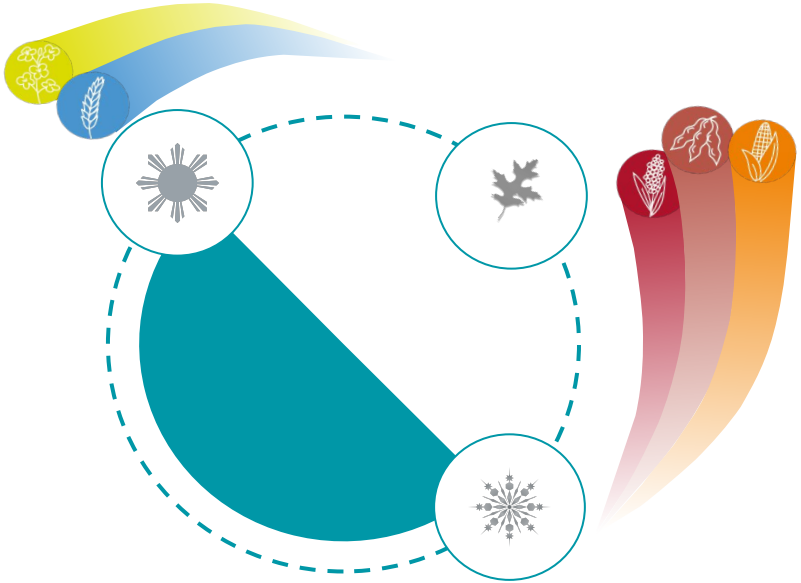
> 15 ans d'expérience

~ 20 appareils

	Critères prédits- routine		Nb données produites	
LABO	> 50	Cal. commerciales / projets coll. / dvpt interne Humidité, protéine, huile, ...	> 1 000 000	↘
EMBARQUE	< 10	Dvpt interne + cal. commerciales Humidité, protéine, huile, ...	> 500 000	↗

En contraintes

1. Saisonnalité



2. Délai d'analyses

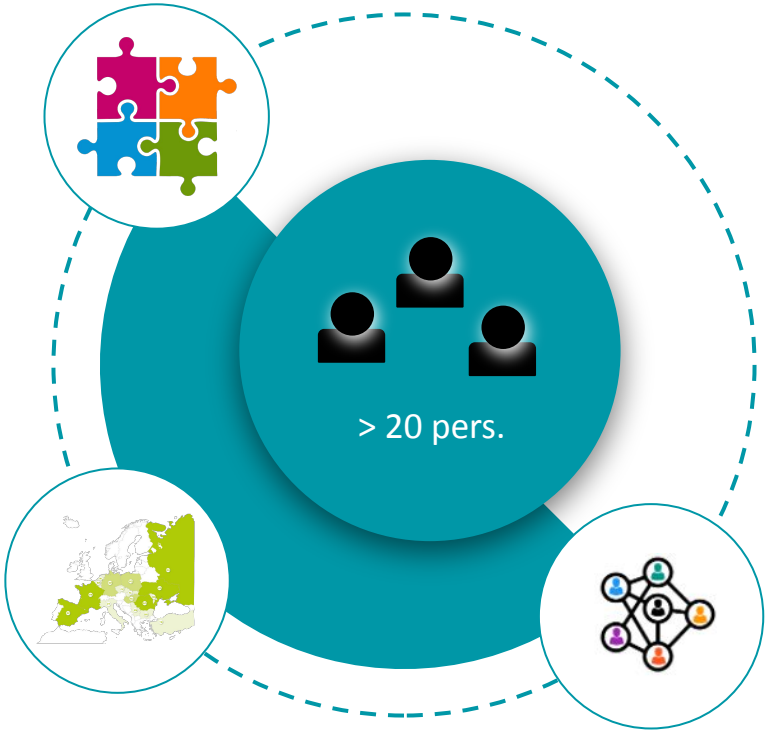
DL (semis, inscription) = figées

2020
Embarqué : 25/06 - 30/07
Labo :+1 s

2021
Embarqué : 01/07 - 09/08
Labo :+3 s

Multi-compétences :

- Machines
- Outils informatiques
- Opération
- Statistiques
- Experts espèces

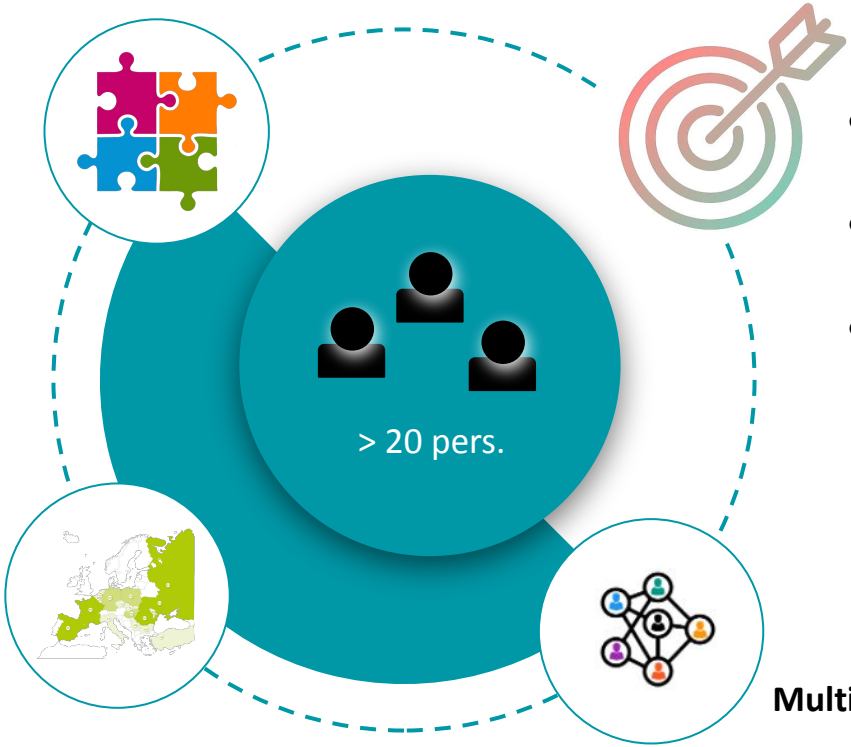


Multi-sites
Collaborateurs &
Machines
National - International
Multilingues

Multi-services

Multi-compétences :

- Machines
- Outils informatiques
- Opération
- Statistiques
- Experts espèces

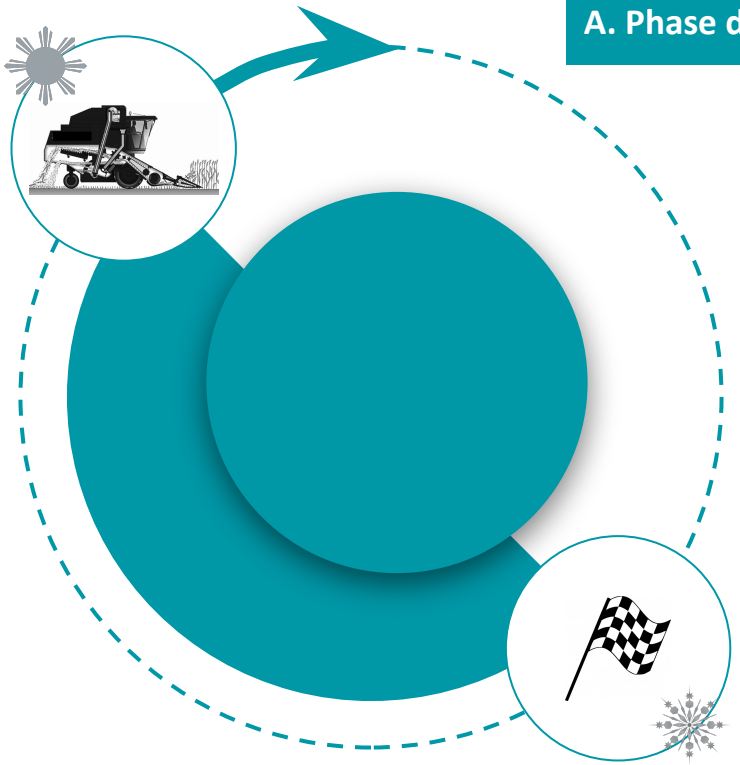


- **Communication**
- **Temps dédié à l'activité**
- **Sécurisation des compétences / expertises**

Multi-services

Multi-sites
Collaborateurs & Machines
National - International
Multilingues

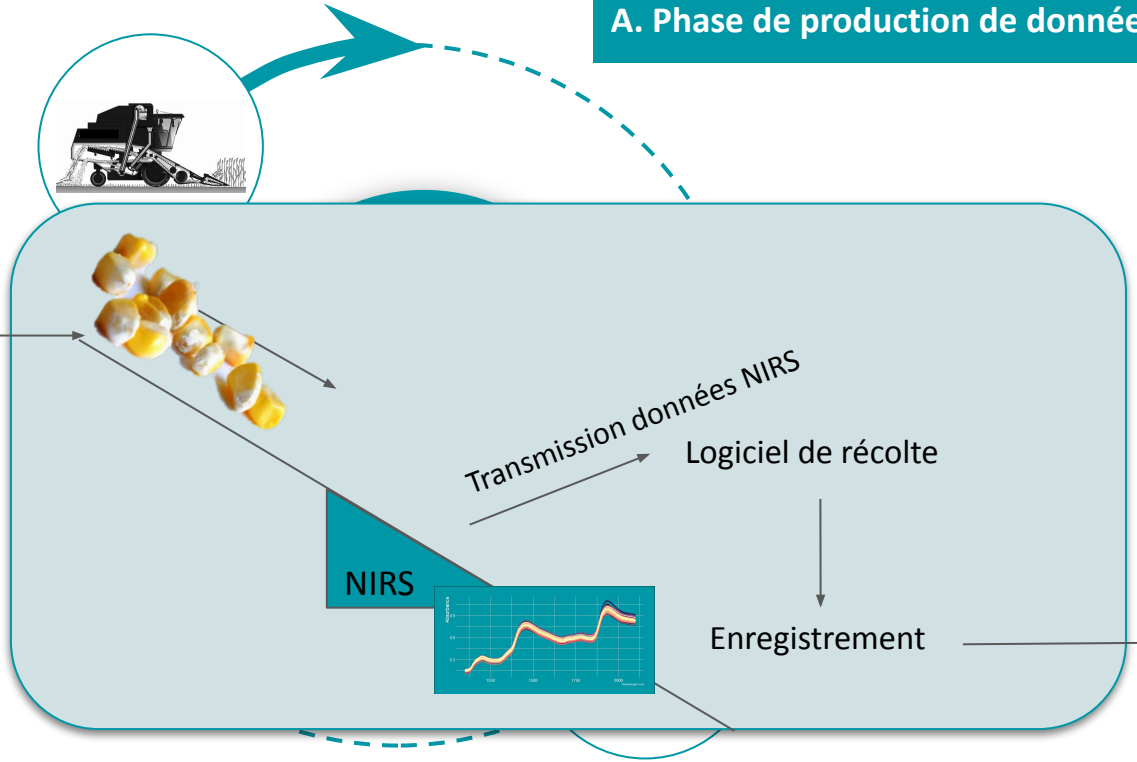
A. Phase de production de données



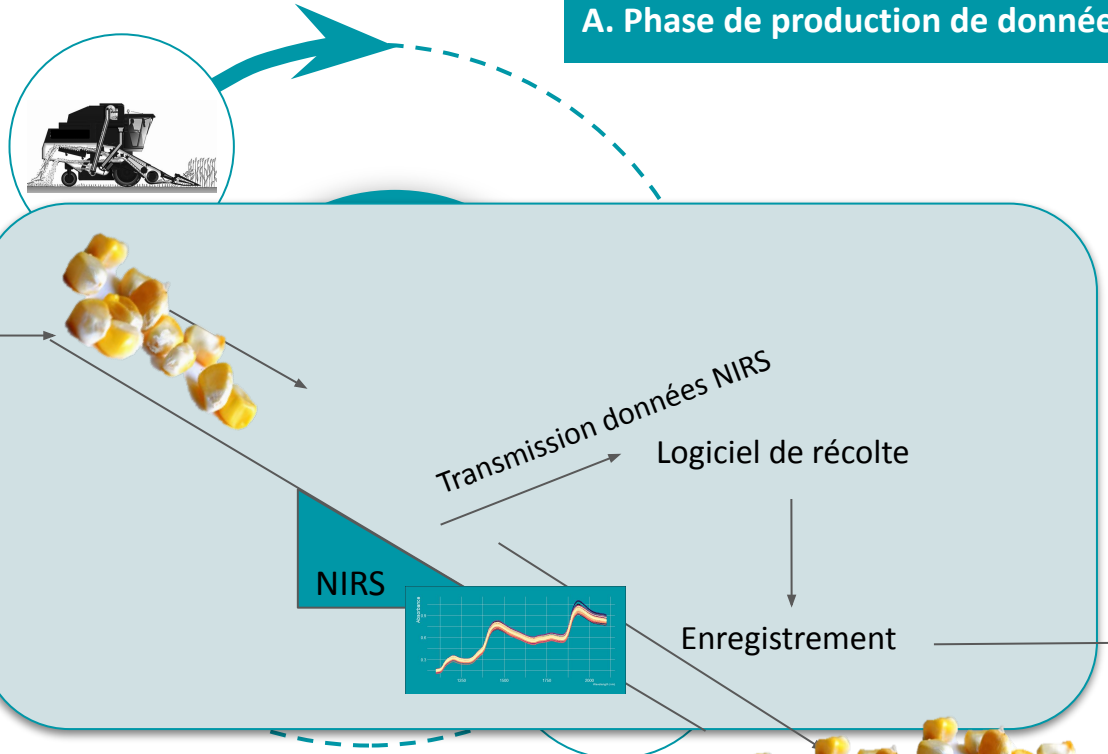
A. Phase de production de données

Récolte parcelle*

**surface qq m²*
même génotype répété x fois
1 champ = x000 parc.



A. Phase de production de données

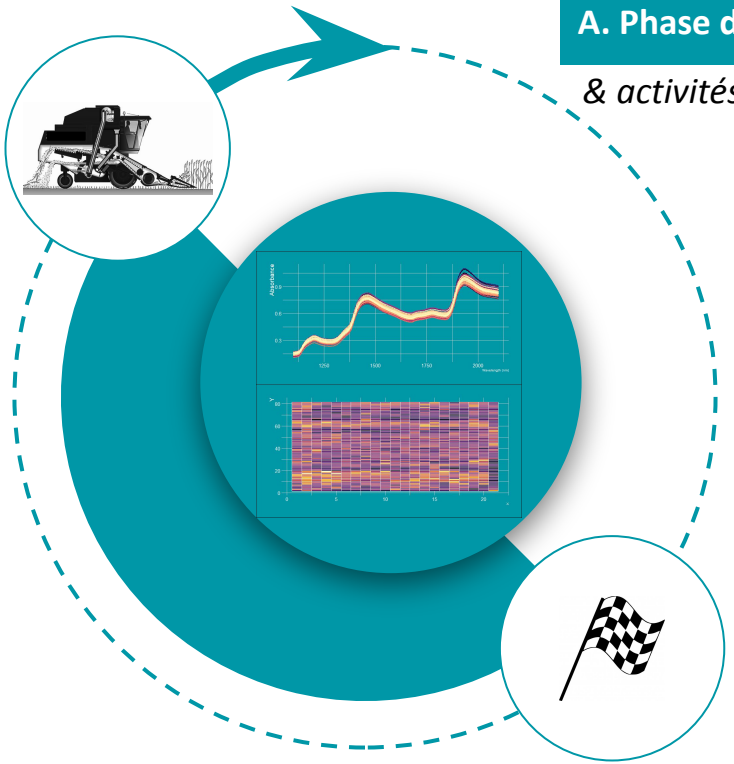


Récolte parcelle*

**surface qq m²*
même génotype répété x fois
1 champ = x000 parc.

Contrainte

Pas de retour en arrière possible !



A. Phase de production de données

& activités “satellites” :

1. **SAV interne**

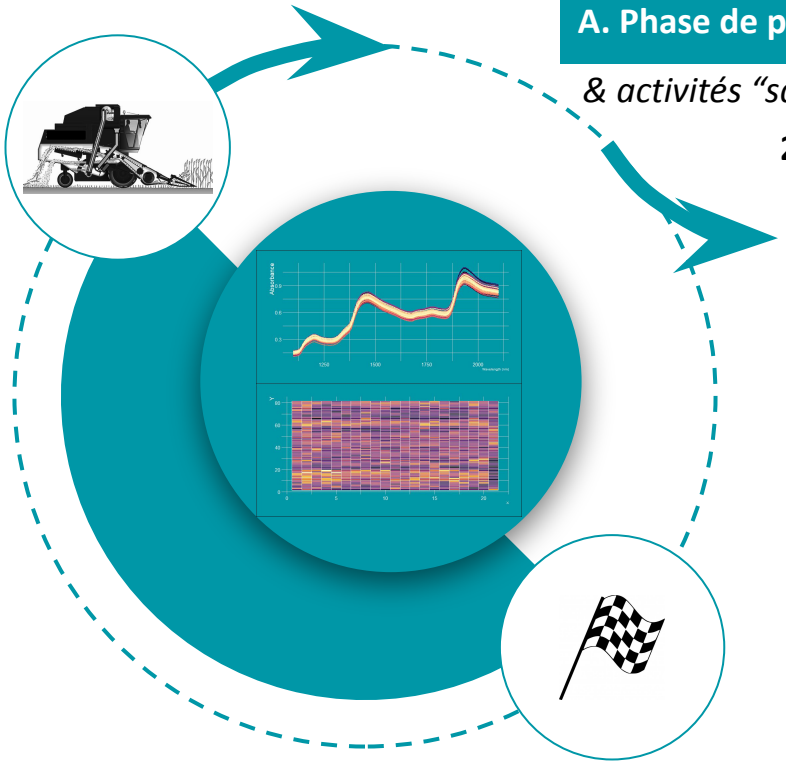
   Machines
   Outils informatiques

   Week-ends + nuits y compris

A. Phase de production de données

& activités “satellites” :

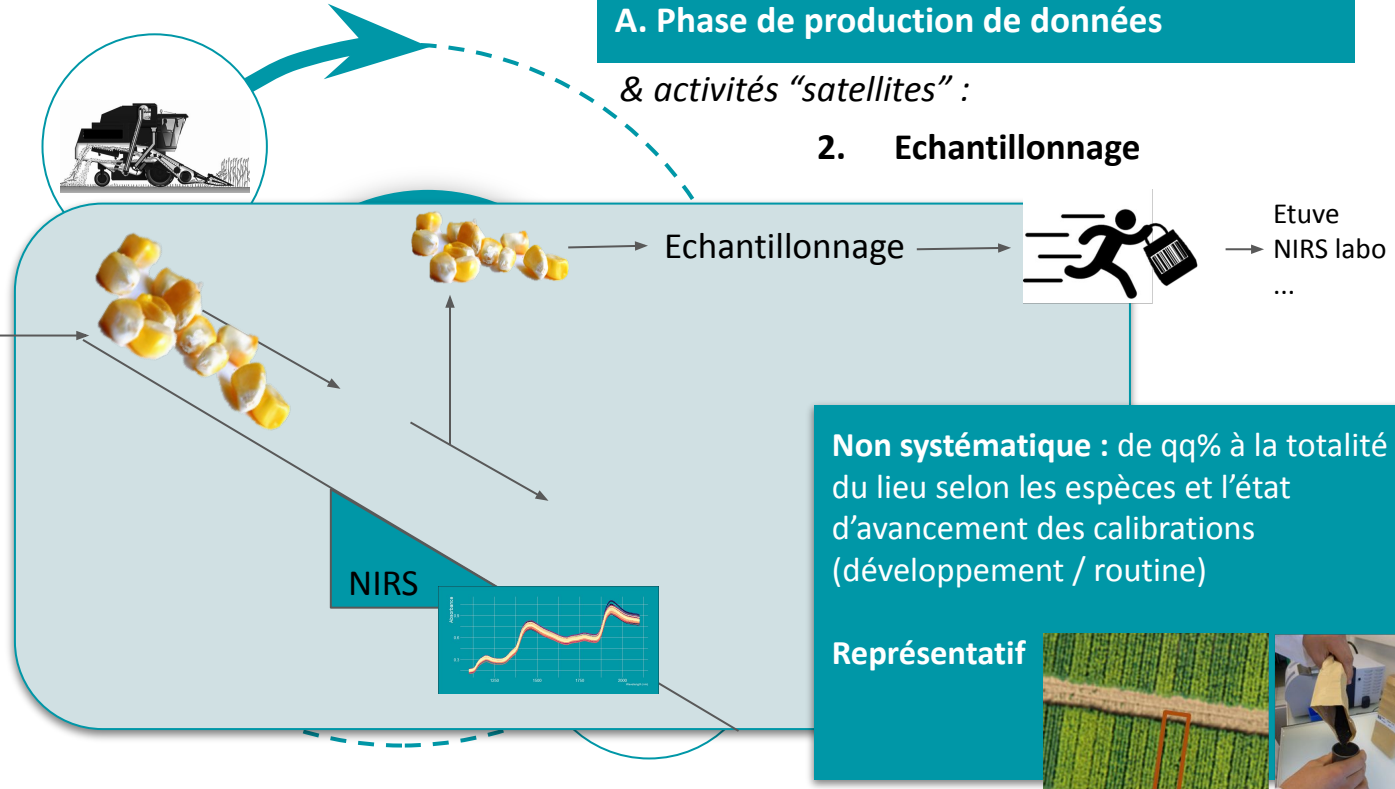
2. Echantillonnage



A. Phase de production de données

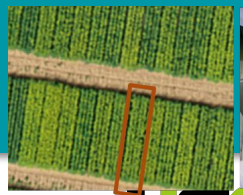
& activités “satellites” :

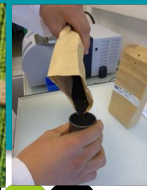
2. Echantillonnage



Non systématique : de qq% à la totalité du lieu selon les espèces et l’état d’avancement des calibrations (développement / routine)

Représentatif





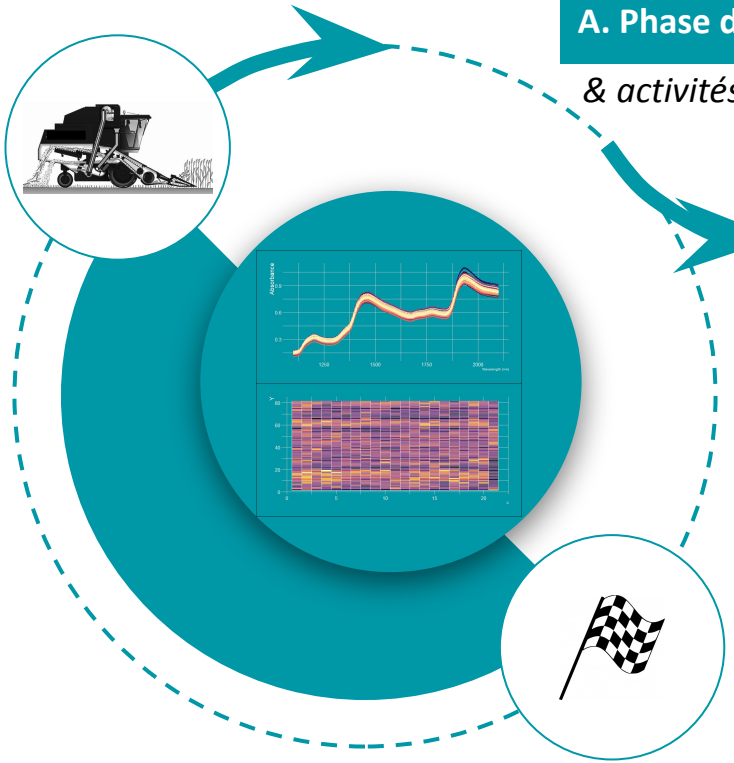
A. Phase de production de données

& activités “satellites” :

2. **Echantillonnage**

↓

3. **Validation données NIRS**

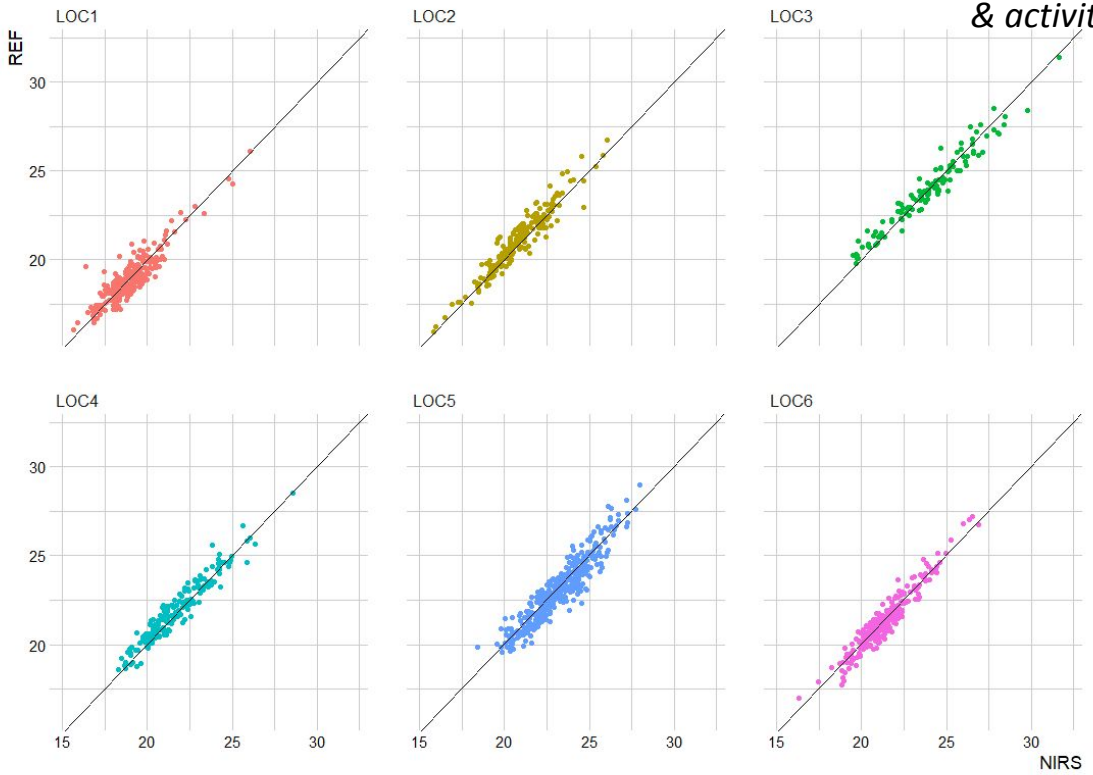


A. Phase de production de données

& activités "satellites" :

2. **Echantillonnage**

3. **Validation données NIRS**



Critères :

- dist Mahal.
- SEP ~SEC
- Correction des biais

A. Phase de production de données

& activités "satellites" :

2. **Echantillonnage**

↓

3. **Validation données NIRS**

Max 1 lieu sur 10 non conforme (SEP > SEC)

Qui est conforme ? NIRS ou REF ?

→ impact très important sur la confiance accordée aux données embarquées de l'ensemble du lieu (pour lesquelles on n'a pas d'échantillonnage)

→ nettoyage des données en vue des mises à jour de calibration

Qq jours pour trouver la solution

A. Phase de production de données

& activités "satellites" :

2. **Echantillonnage**

↓

3. **Validation données NIRS**

Max 1 lieu sur 10 non conforme (SEP > SEC)

Qualité de l'analyse de référence (étuve, NIRS labo, …)

- protocole adapté à la production de données en masse ≠ norme
- Problèmes récurrents : dysfonctionnement étuve, problème de tare, …

SEP :

ε NIRS
ε ref
ε process ref (≠ négligeable)

A. Phase de production de données

& activités "satellites" :

2. **Echantillonnage**
   ↓
3. **Validation données NIRS**

Max 1 lieu sur 10 non conforme (SEP > SEC)

Nouveau NIRS / nouvelles conditions de mesure (+++ humide et ou +++ sec, température) vis-à-vis de la BDD de calibration

SECV ±1.5 // SEP 2021 ±1.5

	Pred. ini	Cal. Lieu2	DOP	TCA
1	1.9	1.7	1.4	1.7
2	2.1	1.3	1.9	2.0
3	1.4	1.3	1.4	1.6
4	1.8	1.3	1.8	1.8
5	2.1	1.9	2.1	2.3

ELSEVIER
Postharvest Biology and Technology
Volume 175, May 2021, 111414

FRUITNIR-GUI: A graphical user interface for correcting external influences in multi-batch near infrared experiments related to fruit quality prediction

Puneet Mishra [a], Jean Michel Roger [b, c], Federico Marini [d], Alessandra Biancolillo [e], Douglas N. Rutledge [f, g]

dynamic orthogonal projection (DOP)
transfer component analysis (TCA)

A. Phase de production de données

& activités "satellites" :

2. **Echantillonnage**

↓

3. **Validation données NIRS**

Max 1 lieu sur 10 non conforme (SEP > SEC)

Dysfonctionnement NIRS (+ rare)

- Dysfonctionnements non bloquants → acquisition de données continue
- Ech. témoins → permet d'anticiper qq problèmes (carte contrôle moins informative qu'en labo)
- Certains cas non anticipés → visualisation a posteriori → Manque d'outils / d'expertise

Max 1 lieu sur 10 non conforme (SEP > SEC)

Dysfonctionnement NIRS (+ rare)

- Dysfonctionnements non bloquants ⟶ acquisition de données continue
- Ech. témoins ⟶ permet d'anticiper qq problèmes (carte contrôle moins informative qu'en labo)
- Certains cas non anticipés ⟶ visualisation a posteriori ⟶ Manque d'outils / d'expertise

A. Phase de production de données

& activités "satellites" :

2. Echantillonnage

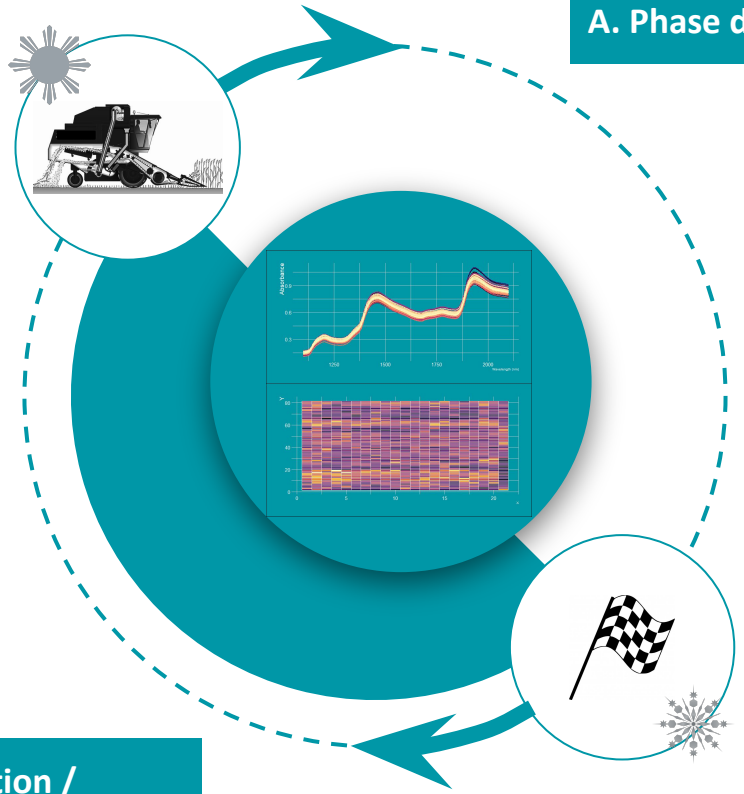
↓

Validation données NIRS

pertinent, i.e. doit couvrir la variabilité lieu + zones éventuelles "à problème"

⟶ Possibilité de travailler les calibrations en cours de route et de sauver des lieux

A. Phase de production de données



B. Phase de bilan / préparation / optimisation

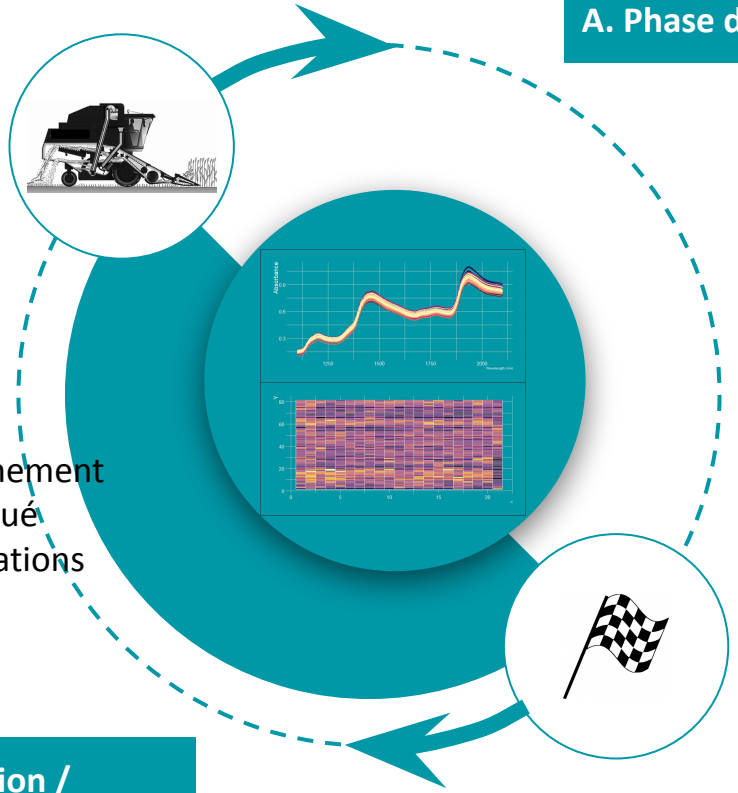
A. Phase de production de données

Opt. logistique / matériel / informatique / calibration

Maintenance matériel

Formation interne / externe

Préparation récoltes
vérification du bon fonctionnement machines + matériel embarqué
implémentation des optimisations



B. Phase de bilan / préparation / optimisation

A. Phase de production de données

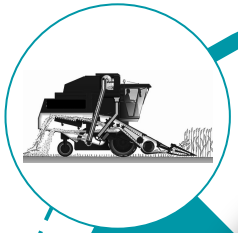
2. **Echantillonnage**
3. **Validation données NIRS**

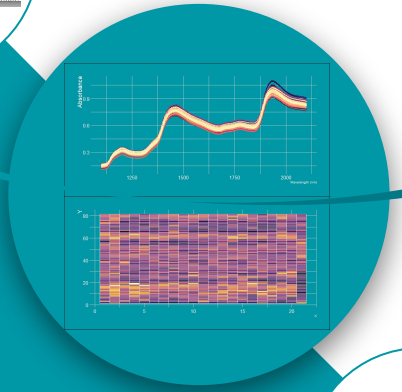
Opt. logistique / matériel / informatique / calibration

Maintenance matériel

Formation interne / externe

Préparation récoltes
vérification du bon fonctionnement machines + matériel embarqué
implémentation des MAJ





B. Phase de bilan / préparation / optimisation

Mise à jour des calibrations

BDD jusqu'à 50 000 à 100 000 spectres + ref / espèce - critère

- après nettoyage
- 5 dernières années de mesure

→

~5-10%

représentatives de la variabilité spectrale / critère / années/ NIRS / lieux

→

PLS

Mise à jour des calibrations

BDD jusqu'à 50 000 à 100 000 spectres + ref / espèce - critère

- après nettoyage
- 5 dernières années de mesure

→

~5-10%

représentatives de la variabilité spectrale / critère / années/ NIRS / lieux

→

PLS

= activité de "routine"

~ 10 jrs/an multi-espèces multi-critères

⟶ automatisation du pipeline de mise à jour de calibrations

Mise à jour des calibrations

BDD jusqu'à 50 000 à 100 000 spectres + ref / espèce - critère

- après nettoyage
- 5 dernières années de mesure

→

~5-10%

représentatives de la variabilité spectrale / critère / années/ NIRS / lieux

→

PLS

Fréquence de maintien des modèles : en cours de réflexion // chaque année = année exceptionnelle !

- 2020 // lieux très humides / très secs
- 2021 // Fusion : nb appareil *2 + nouveaux lieux

Mise à jour des calibrations

BDD jusqu'à 50 000 à 100 000 spectres + ref / espèce - critère

- après nettoyage
- 5 dernières années de mesure

→

~5-10%

représentatives de la variabilité spectrale / critère / années/ NIRS / lieux

→

PLS

Type de modèles :

autres alternatives en cours d'analyse (réseaux de neurones, SVM, …) → amélioration SEP et/ou biais

en se basant sur ce que produit la communauté scientifique

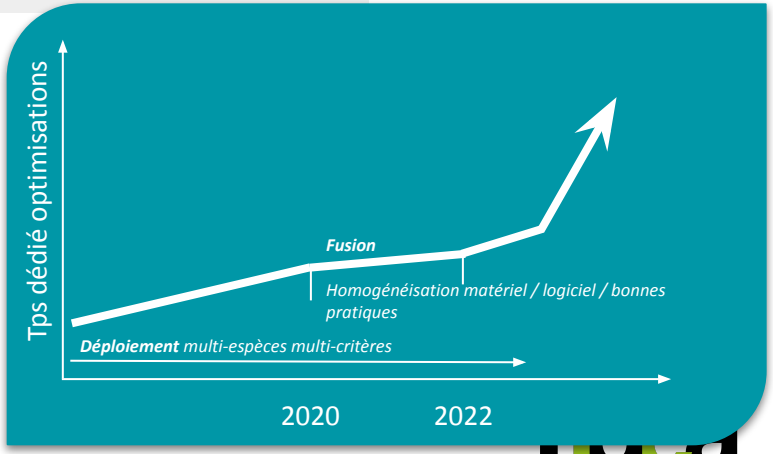
Contrainte : modification du pipeline d'analyses en cours de récolte (prédiction post récolte au plus proche de la récolte...)

FORCES

- Expertise ↗ au fur et à mesure
- Confiance accordée aux données embarquées ↗
- Conséquence : échantillonnage et coûts associés ↘

FAIBLESSES

- Difficulté à anticiper les dysfonctionnements NIRS
- Données : transfert récolte ⟶ stockage / BDD spectrale
- Identification du temps à consacrer à l'optimisation (noyée dans la production de données)



FORCES

- Expertise ↗ au fur et à mesure
- Confiance accordée aux données embarquées ↗
- Conséquence : échantillonnage et coûts associés ↘

FAIBLESSES

- Difficulté à anticiper les dysfonctionnements NIRS
- Données : transfert récolte → stockage / BDD spectrale
- Identification du temps à consacrer à l'optimisation (noyée dans la production de données)

OPPORTUNITES

- vers le tout embarqué ?

MENACES

- Sécurisation des compétences
- Pérennisation de l'optimisation

Merci de votre attention !

www.lidea-seeds.com

Des questions ?

Lidea

FRESH IDEAS FOR AGRICULTURE