

Qualification statistique des données SOMLIT basse-fréquence : une boîte à outils sous R

Pierre-Guy SAURIAU, Léo PEDEMAI, Philippe PINEAU, Hélène AGOGUE, Christophe ARNAUD, Fabien AUBERT, Laureen BEAUGEARD, Martine BRERET, Gaël GUILLOU, Jérôme JOURDE, Nicolas LACHAUSSEE, Benoit LEBRETON, Hélène MONTANIE
L'équipe SOMLIT rochelaise¹

Contexte : Depuis 20 ans, le SNO-SOMLIT (Service d'Observation en Milieu Littoral), aujourd'hui réseau élémentaire de l'IR-ILICO (<https://www.ir-ilico.fr/>), rassemble et coordonne un réseau d'observations de l'environnement marin le long des côtes Manche, Atlantique et Méditerranée. L'acquisition basse-fréquence (15 j.) dans les eaux de surface et à marée haute concerne 15 descripteurs physico-chimiques et biologiques auxquels s'ajoute le pico-nanoplancton.

Cette acquisition se fait dans le cadre de protocoles standardisés sous une démarche qualité élaborée par la cellule qualité SOMLIT (<http://somlit.epoc.u-bordeaux1.fr>) avec une triple approche :

- inter-comparaisons annuelles des pratiques de prélèvements et d'analyses entre laboratoires ;
- qualification analytique des données, descripteur par descripteur, selon une grille de 10 codes qualité ;
- mise en base des données qualifiées garantissant un accès libre aux métadonnées et données.

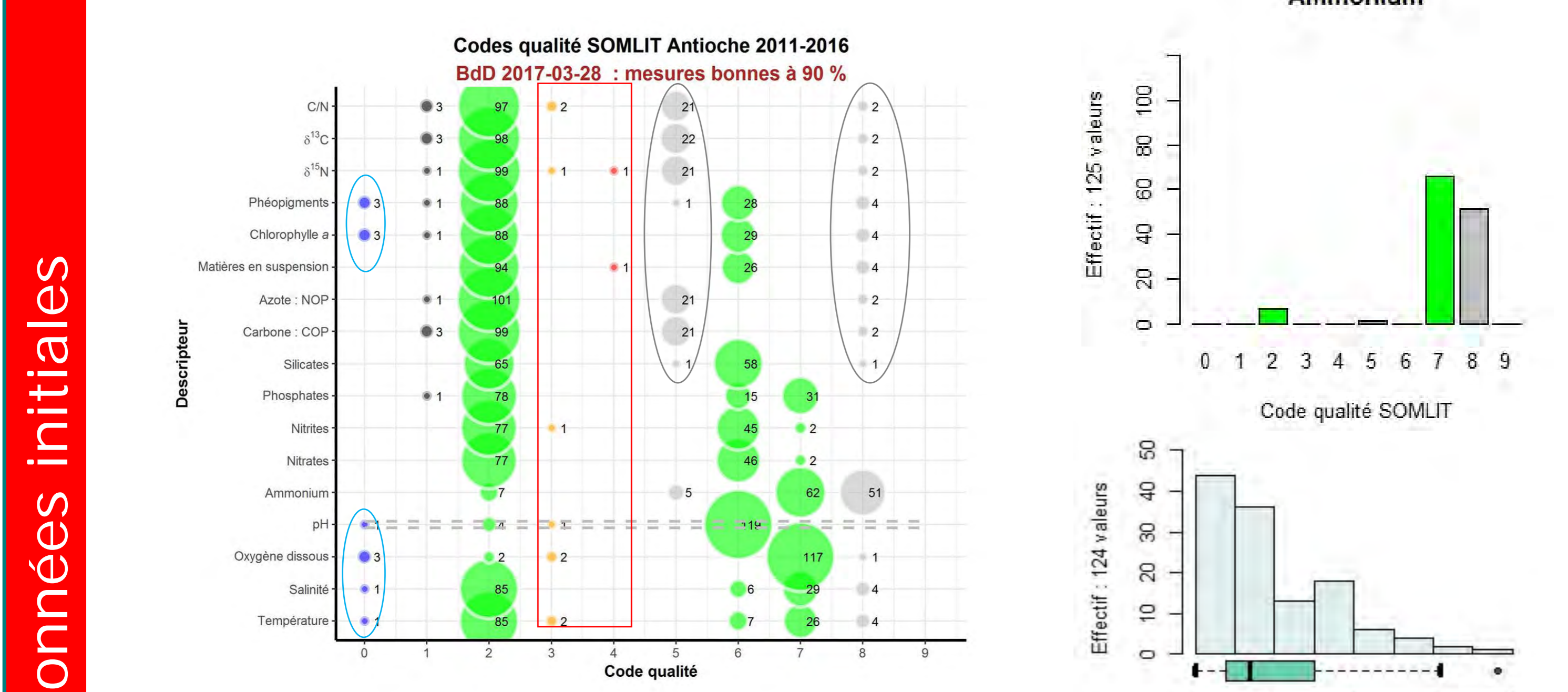


Code Qualité	Description	Code Couleur
0	Donnée en dessous de la limite de détection	●
1	Prélèvement effectué, mais mesure non réalisée	NA
2	Mesure bonne, échantillon non répliqué	●
3	Mesure douteuse	●
4	Mesure mauvaise	●
5	Prélèvement effectué, mais valeur pas encore reportée	NA
6	Mesure bonne (moyenne de plusieurs répliques)	●
7	Mesure bonne (valeur acquise hors protocoles SOMLIT)	●
8	Donnée non qualifiée	●
9	Échantillon non prélevé	NA

Démarche proposée : Ces descripteurs présentent en eux même et/ou entre eux des dépendances fonctionnelles (e.g. cycle O₂ – cycle azote – phytoplancton) qui se traduisent sur le plan statistique par de l'autocorrélation temporelle ou des corrélations significatives entre descripteurs.

Une démarche de qualification statistique des données SOMLIT est proposée, intermédiaire entre la qualification analytique et l'analyse des séries temporelles. Elle s'appuie sur une transcription algorithmique, sous forme de scripts R, de l'étude logique de ces autocorrélations et dépendances multi-descripteurs.

1^{ère} étape : bilan général



Pertinence et justesse des codes qualité sont testés :

- le code qualité 0 (limite de détection) est-il judicieusement utilisé ?
- les codes qualité 3 et 4 (douteux et mauvais) sont-ils justifiés ?
- les codes qualités 5 et 8 (attente de qualification) peuvent-ils être modifiés ?
- le score du nombre de mesures bonnes (ici 90%) est-il juste ou sous-estimé ?

Une pratique homogène et reproductible est suggérée :

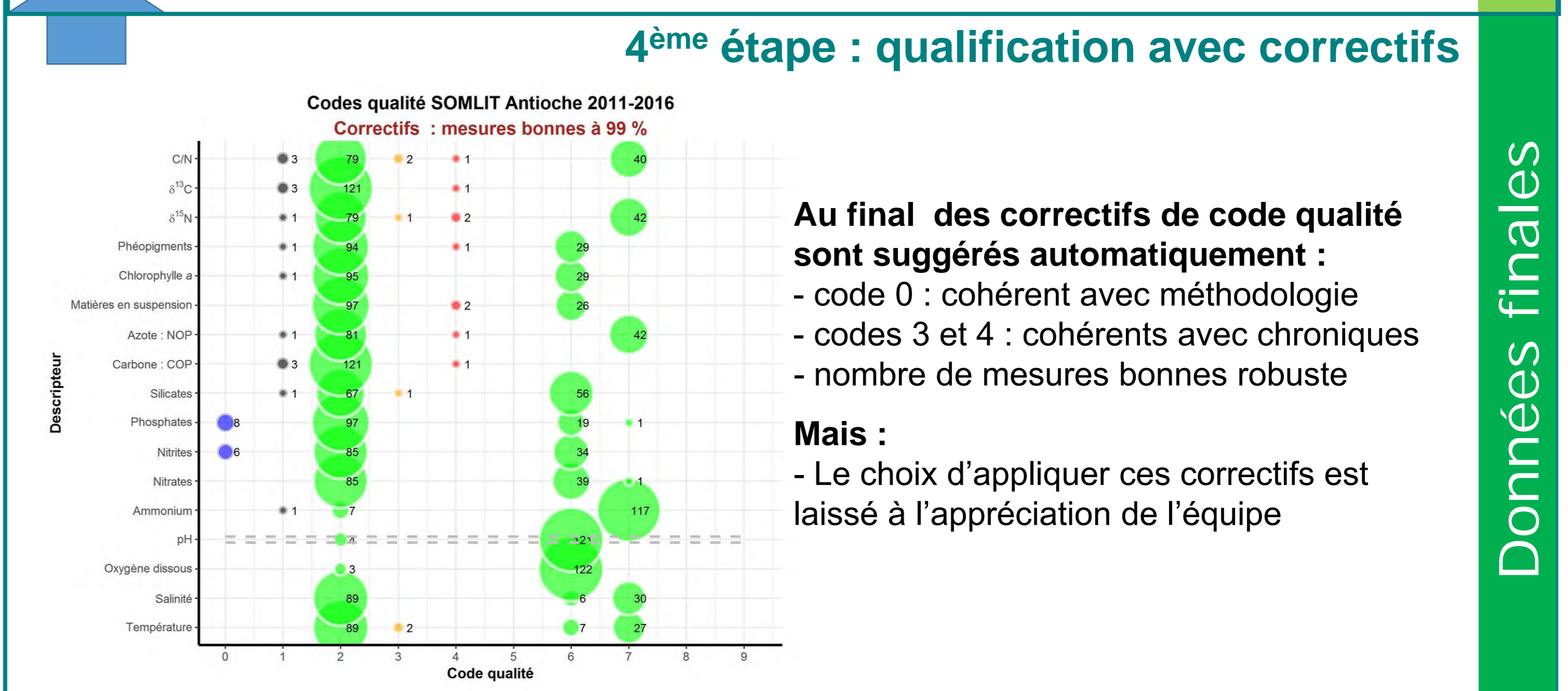
- la pratique de qualification des données, laissée à la discrétion de chaque équipe SOMLIT, peut être homogène et rendue « opérateur-indépendante ».

5^{ème} étape : analyse robuste des séries temporelles

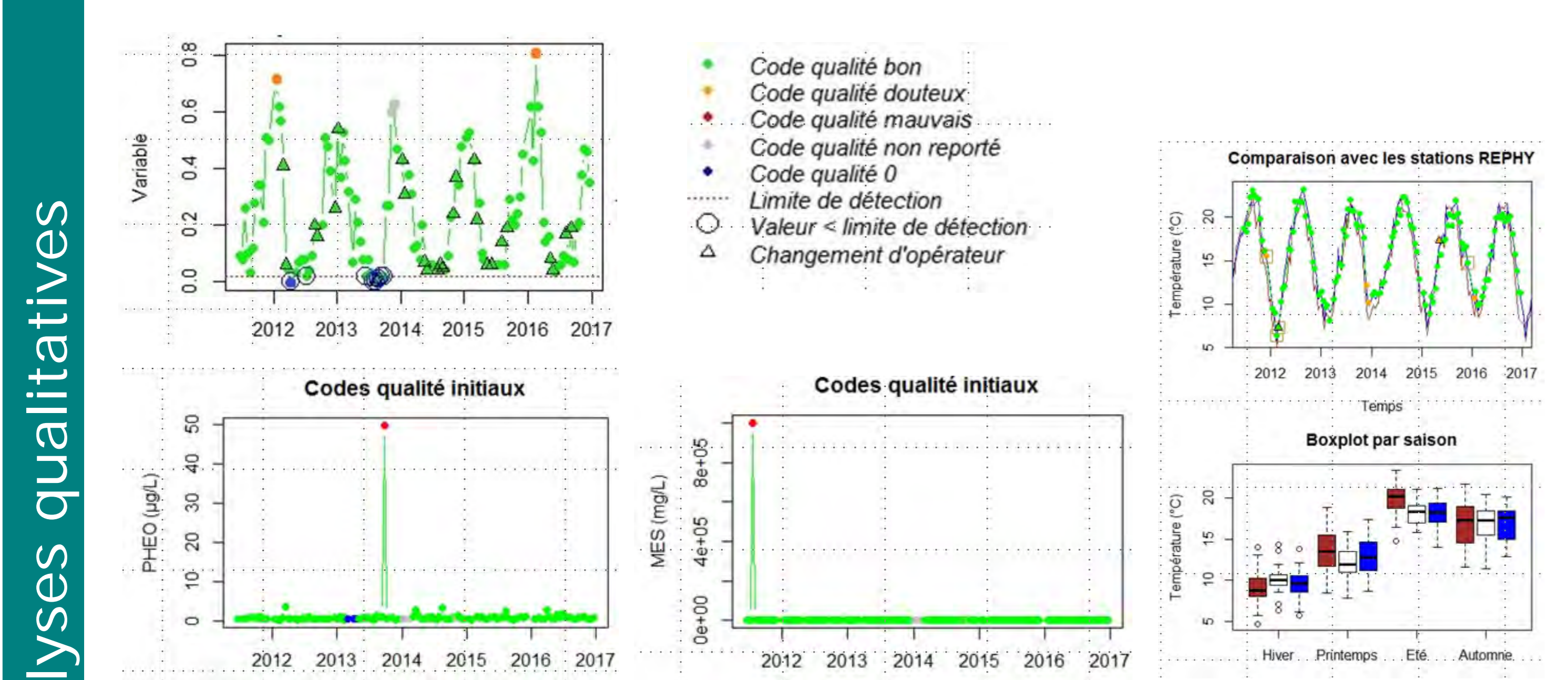
Notre démarche est illustrée sur les données 2011-2016 du site Antioche (La Rochelle). Son application à d'autres sites SOMLIT doit être testée.

Notre démarche renforce la démarche qualité SOMLIT. Elle permet d'écartier du jeu de données des mesures « trop extrêmes » pour lesquelles il y a de bonnes raisons de penser que rien n'explique leur caractère extrême.

Ecartées, ces données pourront alors être interpolées et les séries chronologiques des différents descripteurs analysées et interprétées de façon robuste.



2^{ème} étape : analyse des chroniques



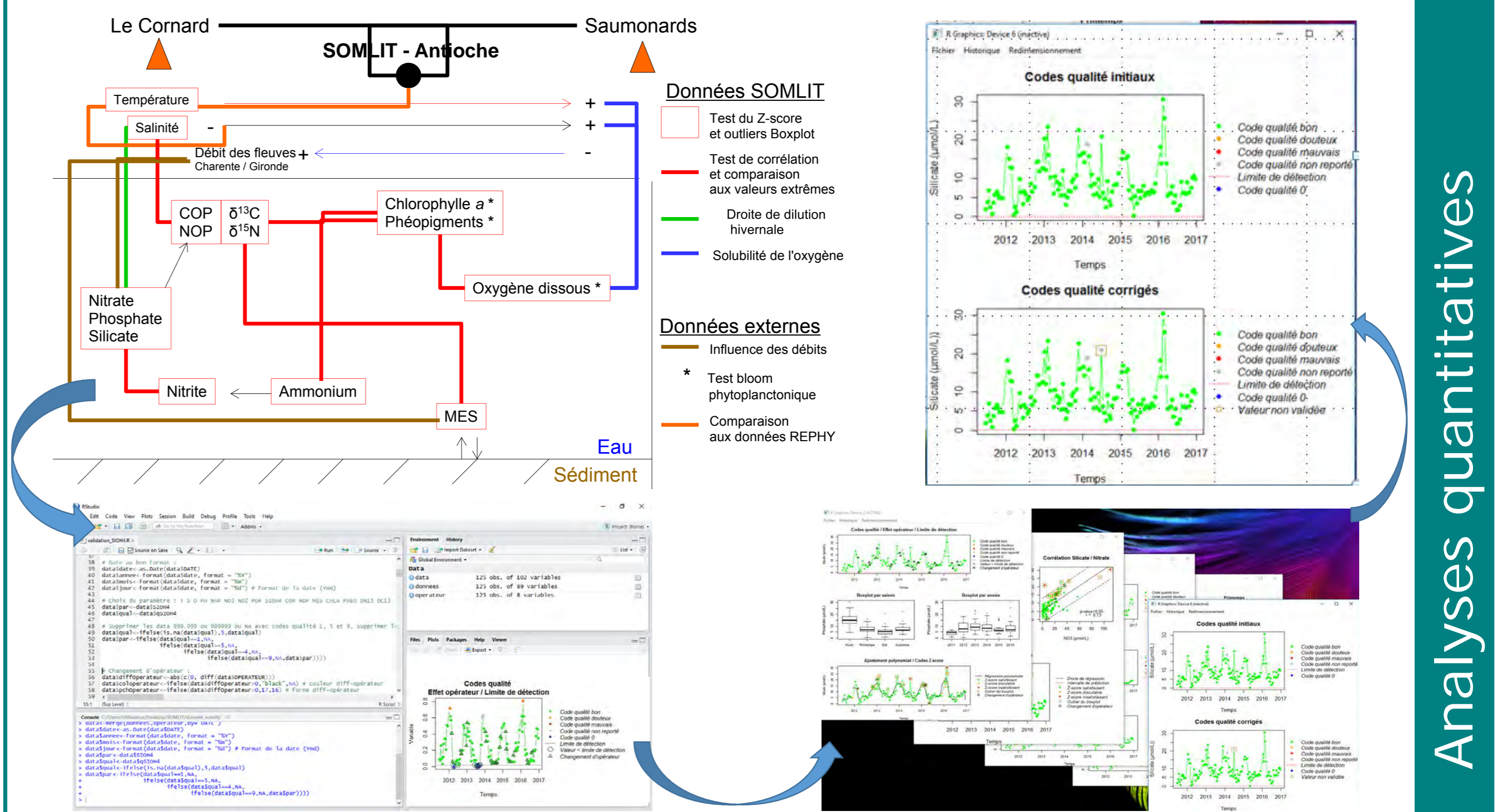
Les chroniques permettent de pointer des valeurs dites « trop extrêmes » :

- valeur trop basse # cohérence des limites de détection ?
- valeur trop haute # valeur extrême expliquée par processus phy/bio connu ?

Besoin d'informations externes à la BdD SOMLIT :

- TESTS sur changements d'opérateur (prélèvement / analyse)
- TESTS sur données météo / blooms phytoplancton / état de la mer / débit fluvial
- TEST sur données physicochimiques indépendantes (T, S, O₂) : REPHY, Coast-HF

3^{ème} étape : analyse des interrelations



Les interrelations entre descripteurs sont testées via des scripts R

Une batterie de sorties graphiques permet un examen pas-à-pas pour chaque chronique d'un descripteur et/ou d'une interrelation entre descripteurs

Objectif final à moyens termes :

interface-web conviviale facile à utiliser
cf. MySOMLIT initié par V. David.

