

Traitement statistique

Restitution de l'étude Aquaref sur le bassin Rhône Méditerranée Corse

Nathalie Guigues et Bénédicte Lepot

1. Estimation des incertitudes de mesure incluant la contribution de l'échantillonnage - surveillance des cours d'eau

Exploitation des données : décomposition de la variance (s^2)

$$s^2_{\text{totale}} = s^2_{\text{mesure}} + s^2_{\text{variabilité milieu}} \quad \text{Plan simple}$$

$$\text{Avec } s^2_{\text{mesure}} = s^2_{\text{prélèvement}} + s^2_{\text{analyse}} \quad \text{Plan déséquilibré}$$

- ❖ Le conditionnement, la conservation et le transport des échantillons jusqu'au laboratoire est incluse dans le prélèvement
- ❖ Evaluation de la composante aléatoire uniquement de l'incertitude de mesure

Composante des incertitudes évaluées :

- Cette étude – U Aquaref
- Laboratoire d'analyse – U Labo

	Echantillonnage		Analyse	
	Dispersion - Fidélité intermédiaire	Justesse	Dispersion - Fidélité intermédiaire	Justesse
U Aquaref	oui		oui	
U Labo			oui	oui

Matériaux de référence certification
Essais interlaboratoires

Essais interlaboratoires échantillonnage

Traitement préliminaire

Calcul du taux de quantification => sélection des substances parmi les 77 analysées

- Échelle de la station – campagnes temporelles
- Echelle du bassin – 35 stations de la campagne spatiale / 15 stations de la campagne temporelle
 - Paramètres physico-chimiques majeurs : entre 81 % et 100 %
 - Métaux : entre 53 % et 100 %
 - Micropolluants organiques : entre 3 % et 100 %. Seule la carbendazine n'a pas été détectée

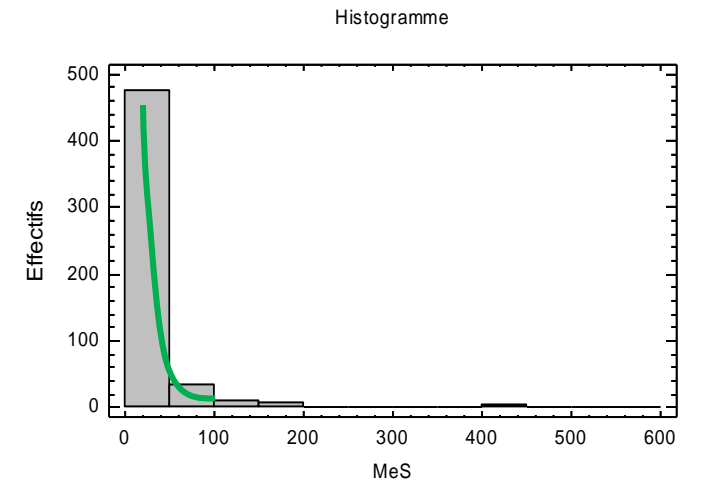
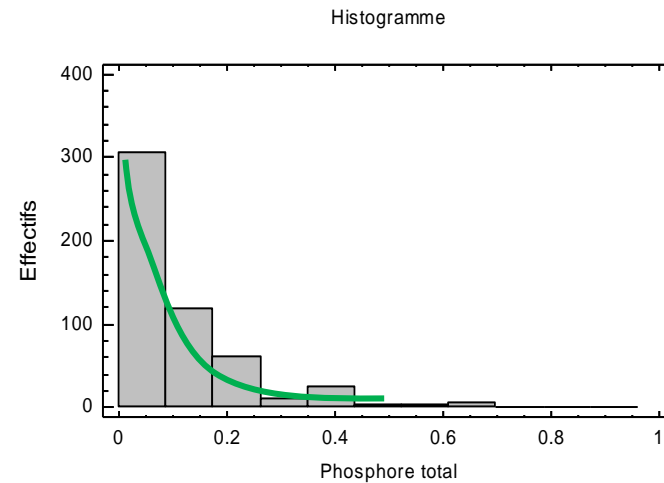
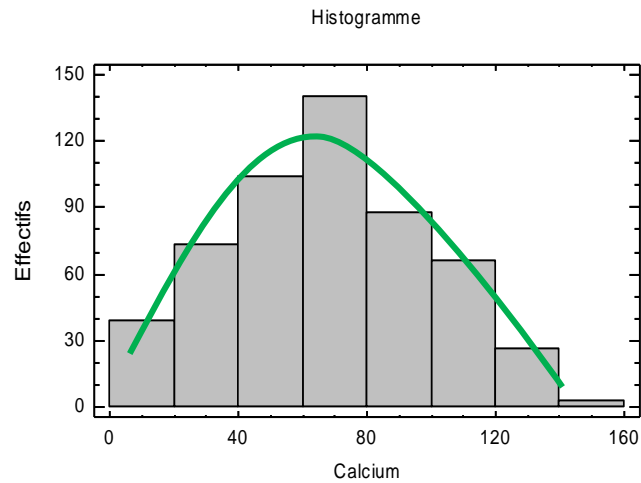
Critère retenu pour réaliser le traitement statistique :

- taux de quantification > 40%
 - 52 paramètres – campagnes temporelles
 - 53 paramètres - campagne spatiale

Traitement préliminaire

Histogrammes de distribution

=> Classement par typologie

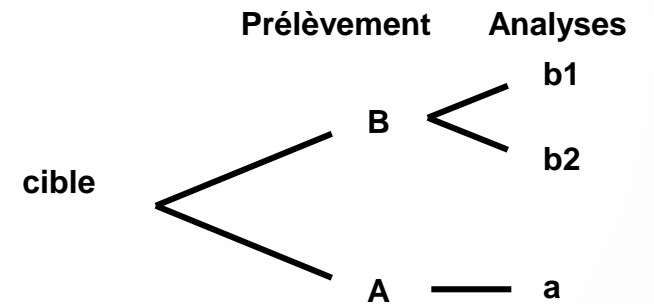


Traitement préliminaire

Remplacement des données $< LQ$ par $LQ/2$

Identification des valeurs aberrantes

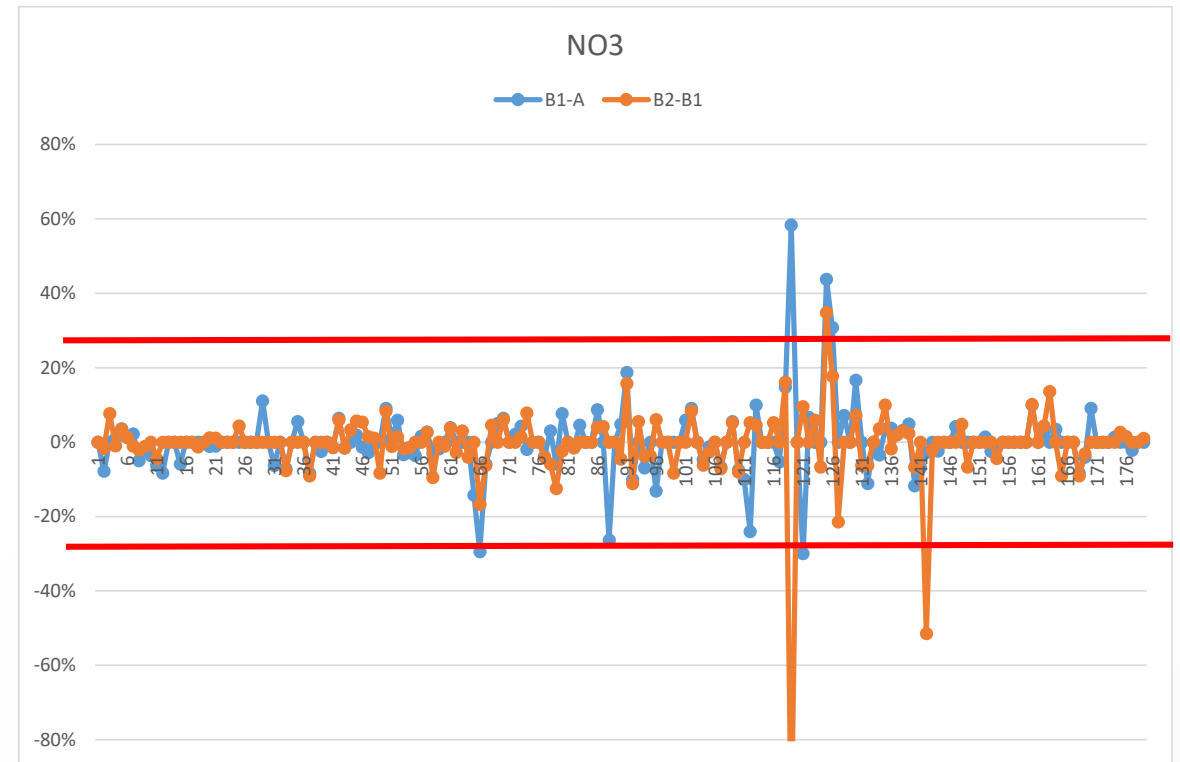
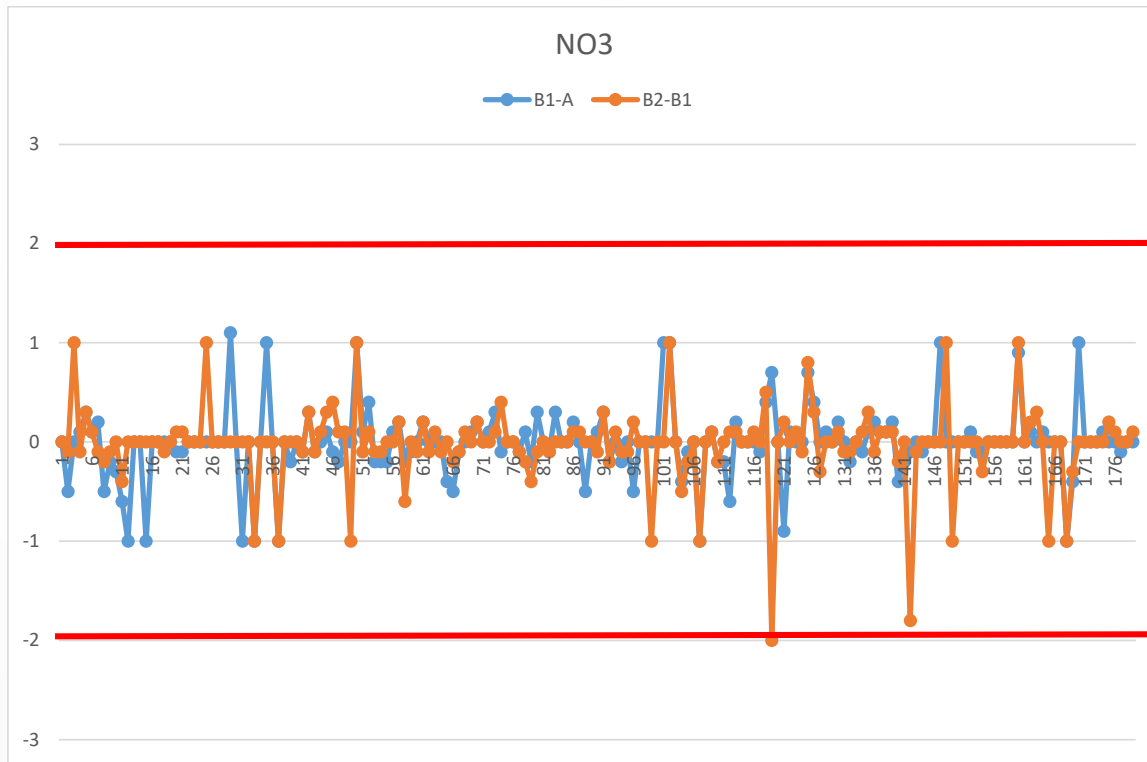
- Différences $a-b1$ et $b1-b2$ en valeur et en relatif (%)
- Définition de seuils de détermination des points aberrants
 - Différence en valeur $> 3 \times LQ$
 - Et différence relative $>$ à U analytique ou 50% (la plus grande)



Exemple des nitrates

- Seuils de détermination des points aberrants : $> 2 \text{ mg/L-NO}_3$ ($3 \times \text{LQ}$) et 25% (U Labo)

 0 / 179 échantillons éliminés

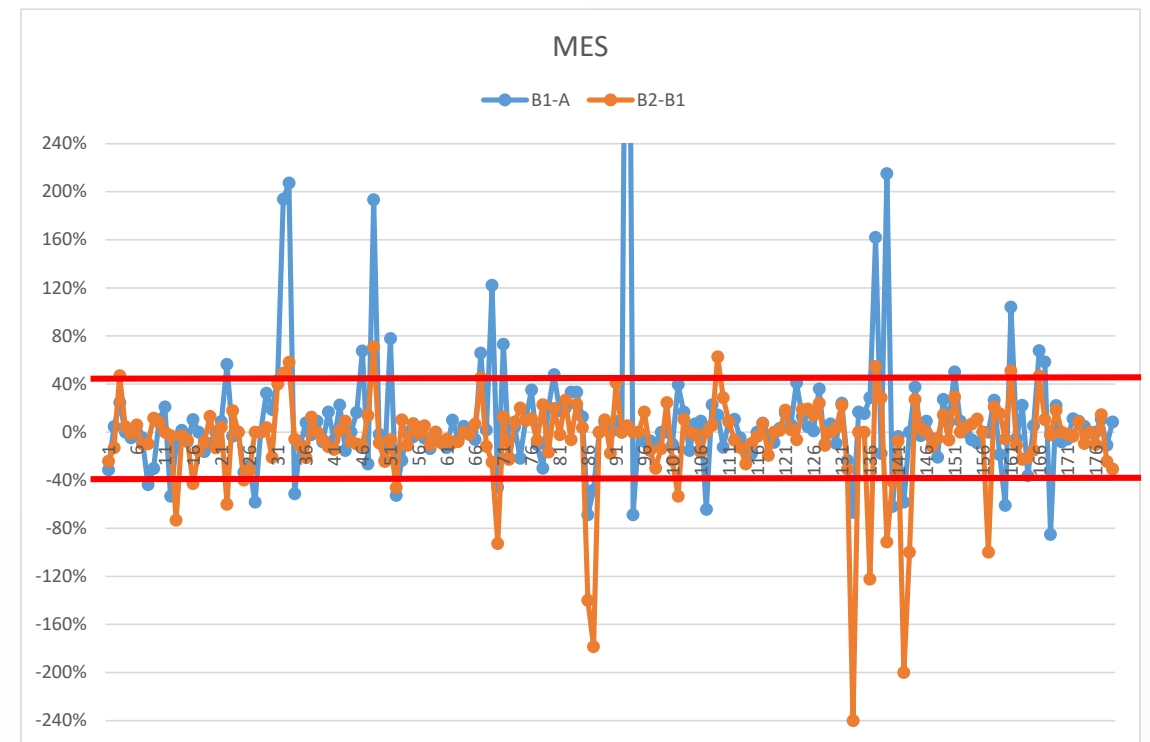
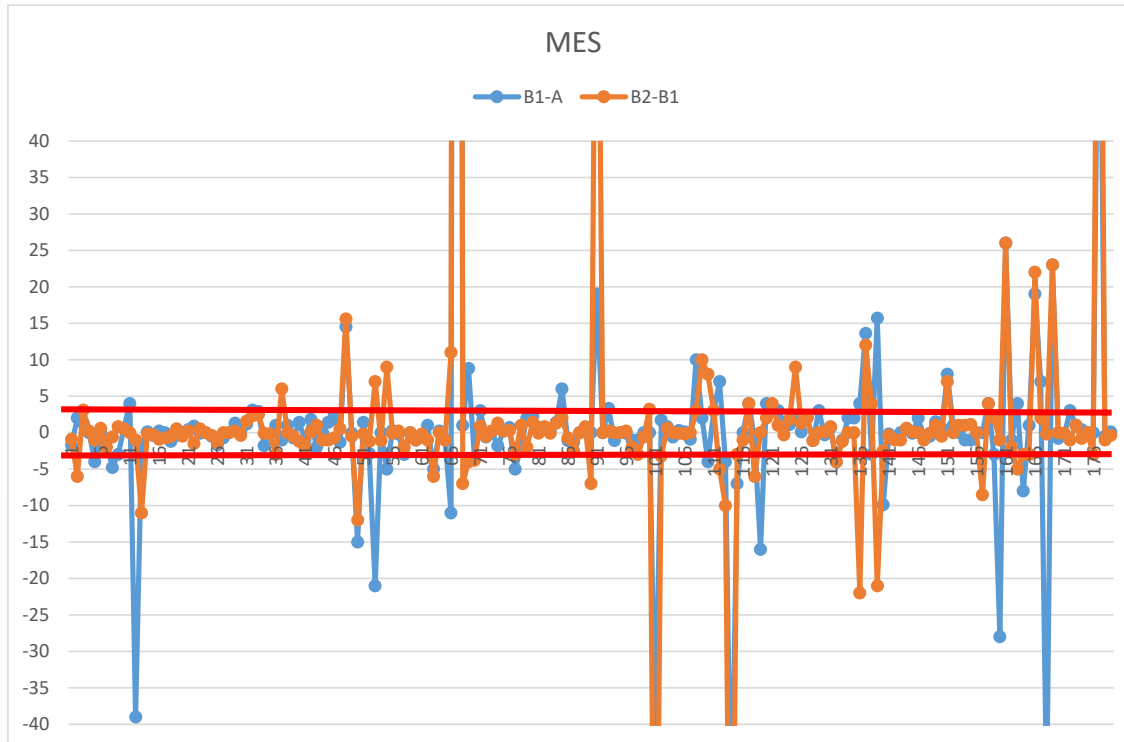


Exemple des MES

- Seuils de détermination des points aberrants : $> 3 \text{ mg/L}$ ($3 \times \text{LQ}$) et 40% (U Labo)



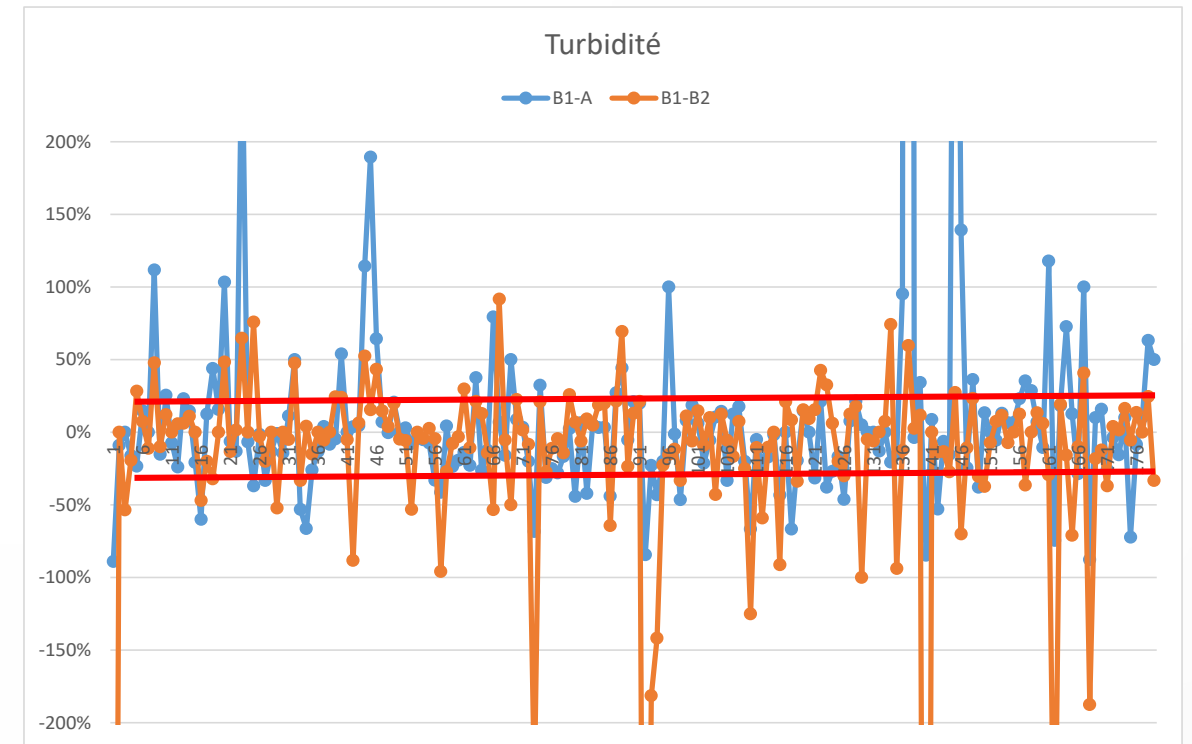
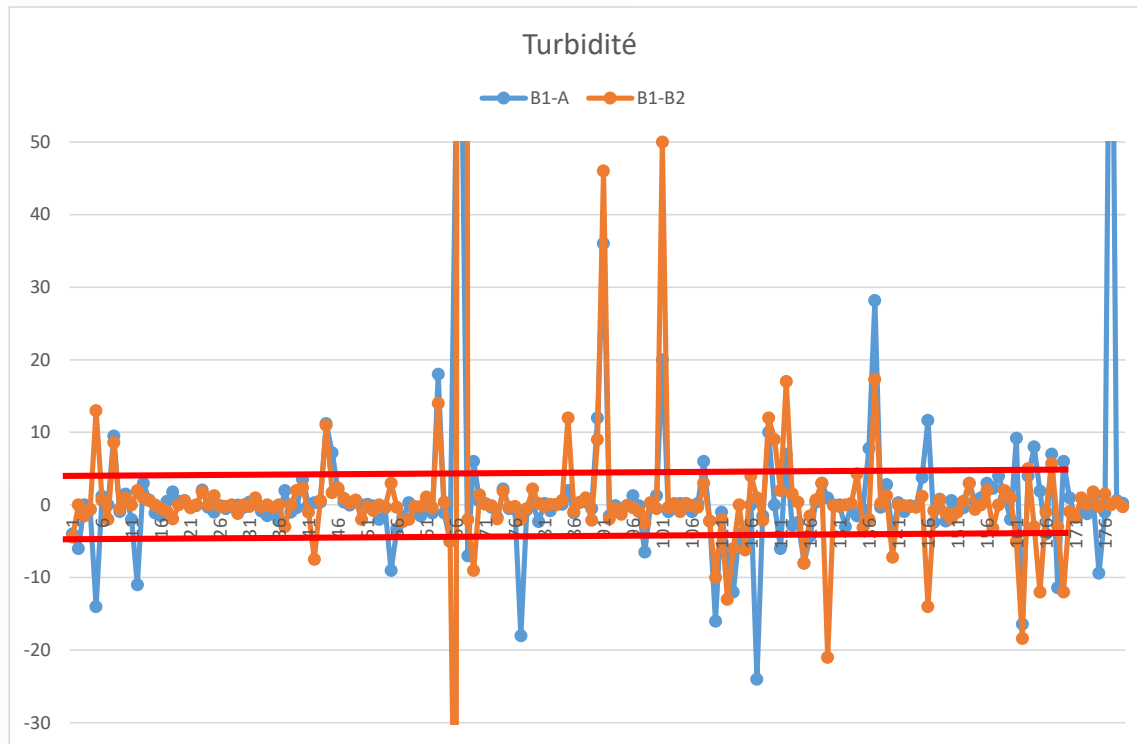
7 / 179 échantillons éliminés



Exemple de la turbidité

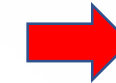
- Seuils de détermination des points aberrants : écart > 2 NTU (arbitraire) et 25% (U Labo)

 23 / 179 échantillons éliminés

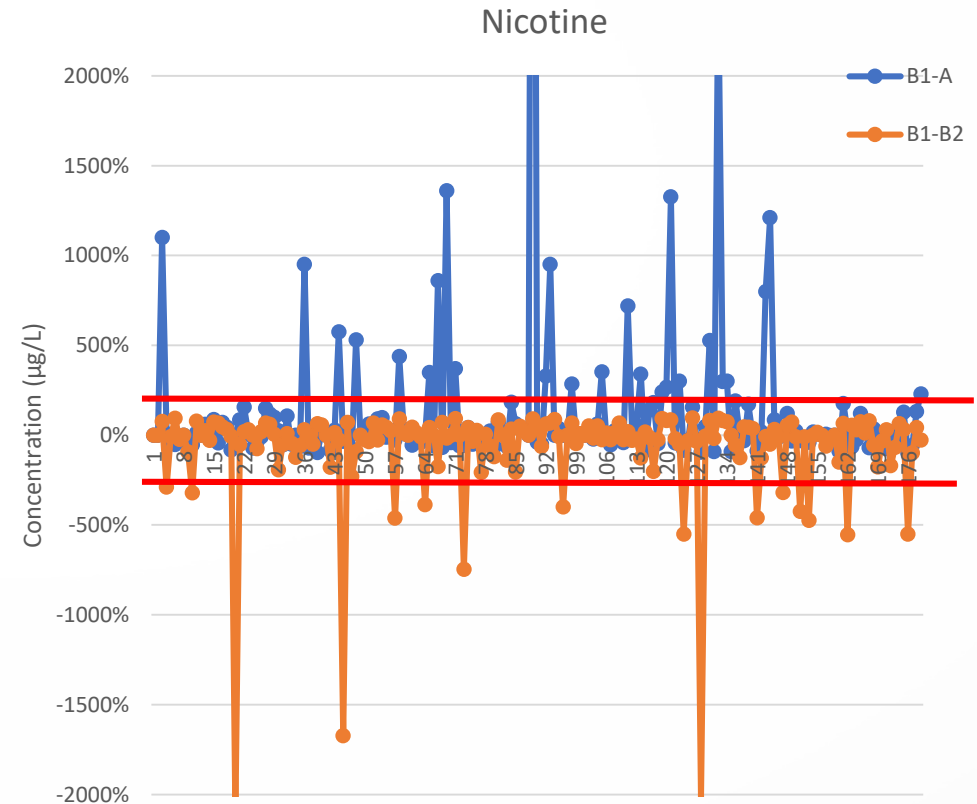
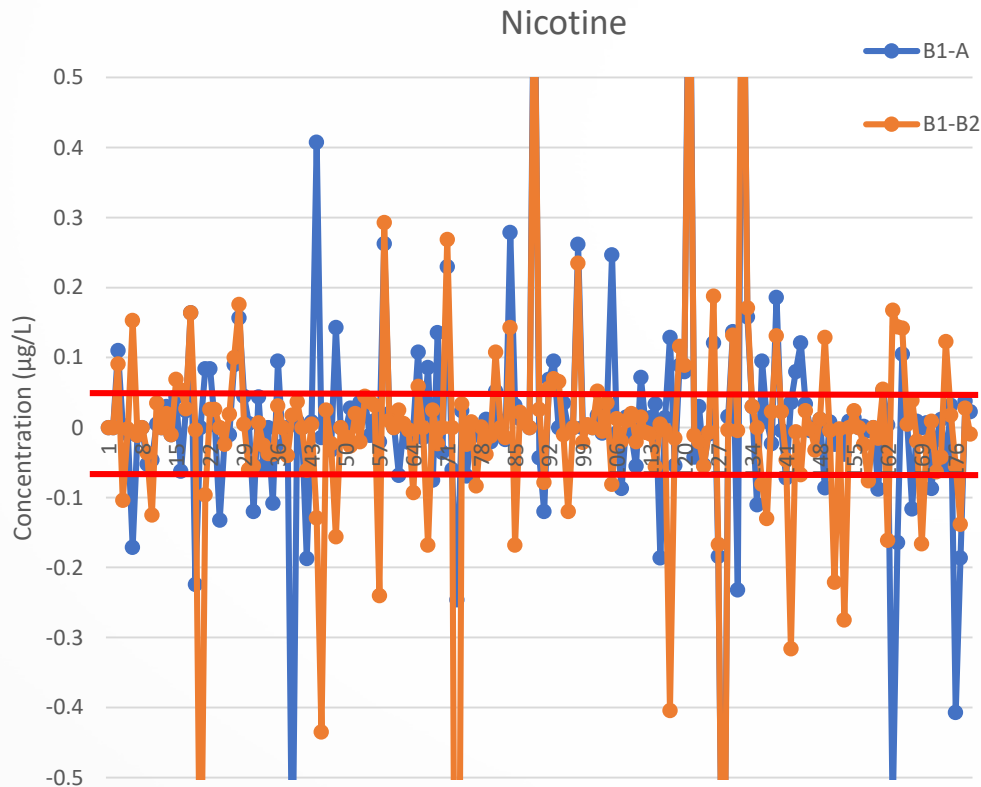


Exemple de la nicotine

- Seuils de détermination des points aberrants : écart > 0.06 µg/L (3 x LQ) et 50%



47 / 179 échantillons éliminés



Nombre de données aberrantes > 10 (plus de 5% des données)

Substances	Nombre de données éliminées	% de données éliminées
Nicotine	47	26%
Turbidité	23	13%
nButyl Phtalate	20	11%
Caféine	17	9.5%
Fer	13	7.3%
Cotinine	10	5.6%
EDTA	10	5.6%

Traitement statistique : analyse de variance (logiciel RANOVA 3)

Classique (ANOVA)

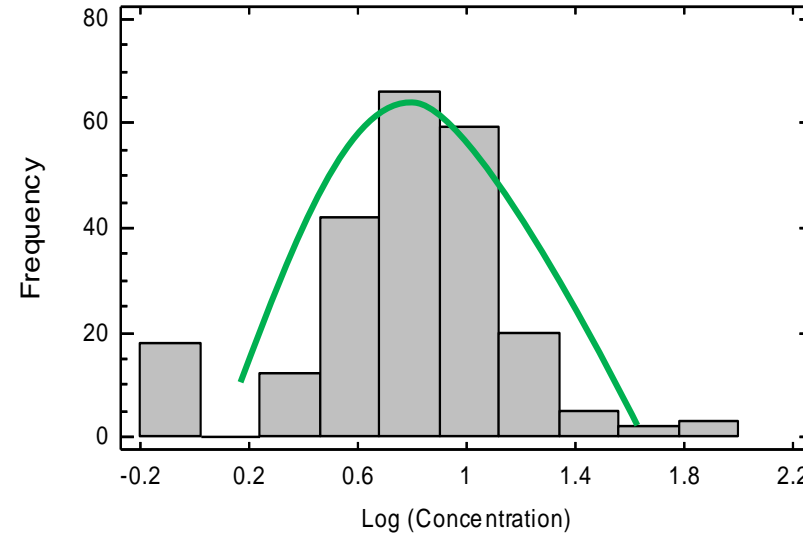
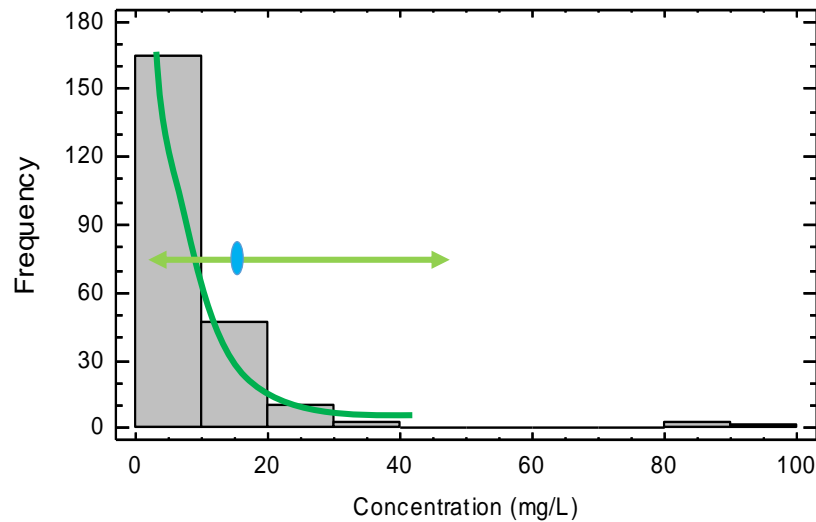
- hypothèse de distribution des données selon une loi (pseudo) normale
- homogénéité des variances et élimination des valeurs aberrantes

Robuste (RANOVA)

- pas d'hypothèse sur la distribution des données
- homogénéité des variances non indispensable et données extrêmes incluses mais pondérées

Alternative pour des incertitudes élevées : ANOVA classique sur les données log-transformées et estimation d'un facteur d'incertitude (FU)

Facteur d'incertitude



$$UCL = x * FU$$

$$LCL = x / FU$$

x

- ⇒ FU permet de rendre compte de l'asymétrie dans la distribution des données
- ⇒ Pour $U < 20\%$, $FU < 1.7$ (asymétrie faible)
- ⇒ Pour $U > 20\%$ $FU > 1.7$ traduit une asymétrie moyenne à forte, introduisant une erreur importante sur l'estimation de l'incertitude U par ANOVA classique ou robuste

Stratégie pour estimer U ou FU

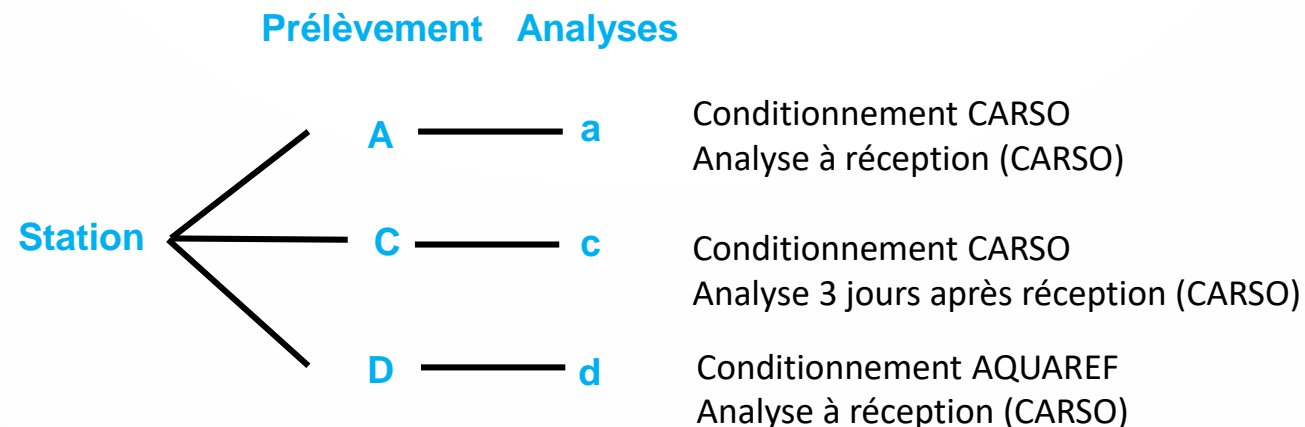
ANOVA classique	ANOVA classique ou robuste selon les stations	ANOVA robuste	ANOVA robuste ou FU selon les stations	FU
Ca, Mg, Na, Cl, SO4	NO3	K, COT, SiO2	PO4, Ptot	NH4, NO2, MES, turbidité
		As, B, Ba, Co, Li, Ni, V, U	Cu, Fe, Mn	Al, Se, Ti, Zn
		Carbamazepine, metolachlore OXA	Irbesartan, Oxazepam	Toutes les autres substances organiques

2. Etude sur l'influence de la température sur la stabilité des échantillons

Traitement statistique : analyse de variance (logiciel RANOVA3)

Estimation de U / FU (RANOVA 3)

- Même stratégie pour estimer U / FU par paramètre
- Estimation U / FU avec les données A et D
- Estimation U / FU avec les données A, D et C
- Calcul du rapport $R = U_{ADC} / U_{AD}$ ou FU_{ADC} / FU_{AD}



Traitement statistique : analyse de variance (logiciel RANOVA3)

Impact du transport (délai et T)

- Identification des substances tels que $R \geq 1,5$ et $R \leq 0,6$
- Impact du délai et température est significatif pour ces substances si U_{ADC} / FU_{ADC} est supérieur aux U / FU estimés lors des campagnes spatiales, temporelles et à l'échelle de la station (U_{max} / FU_{max}).