

# Sites, données et méthodologie

## 2.1 Caractéristiques générales des sites et données

La base de données inclut 70 stations pour les MES, 47 stations pour les SDT, 20 stations pour les nutriments dissous (nitrate, ammoniacque, orthophosphate) et 14 stations pour les nutriments totaux (Phosphore total, azote Kjeldahl total), totalisant plus de 2000 années stations. Cette base de données est la plus importante utilisée à ce jour pour déterminer les incertitudes sur les flux.

### 2.1.1 Stations matières en suspension

La base de données est constituée à plus de 80% de rivières américaines (Figure 2-1). Ces stations couvrent une gamme très large de conditions hydrologiques, climatique, lithologique, d'occupation du sol. Les caractéristiques de chaque station sont détaillées au tableau 2-1 où les stations sont classées par valeurs croissantes de l'indicateur de durée des flux de matière,  $M_{2\%}$ .



Figure 2-1 : Localisation des stations MES issues de la base de données USGS

La base de données MES est composée de stations disposant de 3 à 42 ans de données journalières de concentrations et débits. Ces stations regroupent des bassins versant de taille moyenne à très grande (500 km<sup>2</sup> à 1 800 000 km<sup>2</sup>, médiane = 9 000 km<sup>2</sup>) avec des caractéristiques de variabilité hydrologique et sédimentologique contrastées. Les débits spécifiques varient de 0.05 l/s/km<sup>2</sup>, pour des rivières à régime semi aride ou aride, jusqu'à 40 l/s/km<sup>2</sup> pour certaines rivières de la côte

Pacifique. Les gammes de concentrations sont également très grandes pouvant dépasser les 20 000 mg/l pour des rivières californiennes (Tramblay et al., 2010). Les flux spécifiques de MES varient de 3 t/an/km<sup>2</sup> à 7 300 t/an/km<sup>2</sup>.

Ceci se traduit par des gammes de variabilités des indicateurs de durée des flux d'eau ( $W_{2\%}$ ) et de matière ( $M_{2\%}$ ) très élevées. Les valeurs de  $M_{2\%}$  peuvent même atteindre 98%, c'est-à-dire que 98% du flux de matière (en année moyenne) est transporté en une semaine.

Différents types de relation concentrations-débits sont également représentés avec une très large majorité du type **c-C** (55 stations) puis **s-C** (13 stations) et enfin une station pour chaque type **s-S** et **c-S**

Cette base de données en MES offre donc un large panel de stations qui permet d'étudier les incertitudes sur les flux pour des rivières avec une variabilité des concentrations et des débits faible comme la Seine, jusqu'à des rivières « éphémères » de milieu aride ou semi aride, comme certaines rivières californiennes.

Tableau 2-1 : Caractéristiques générales des stations MES

Rivières	Stations	Code	BV (km <sup>2</sup> )	An	Q* l/s/km <sup>2</sup>	Y* (t/an/km <sup>2</sup> )	C* (mg/l)	M <sub>2%</sub> (%)	W <sub>2%</sub> (%)	b <sub>50sup</sub>	b	Typologie Relation C-Q
Missouri	Culbertson, MT	mis	237030	5	1.52	21	290	16	6	1.36	0.92	c-C
Rhin	Maxau, Germany	rhin	50196	19	25.6	64	53	16	5	1.13	0.98	c-C
Seine	Poses, France	poses	65000	3	7.4	24	68	16	7	1.06	0.93	c-C
Oise	Mery, France	oise	16972	10	7.2	23	67	17	8	0.67	0.82	c-C
Charente	France	charente	7600	4	9	38	89	18	11	-0.15	0.06	c-S
Tennessee	Paducah, KY	ten1	104073	6	15.1	162	225	21	10	0.69	0.88	c-C
Marne	Neuilly, France	mame	12710	10	9.5	45	99	22	8	1.07	0.87	c-C
Sacramento	Freeport, CA	sac	65403	9	11	78	149	23	7	0.91	1.06	c-C
Des Moines	Saylorville, IA	dmoav	15122	7	5.7	257	947	23	12	0.54	0.71	c-C
Mississippi	Saint Louis, MO	missip	1805221	42	21.1	13	13	23	9	1.2	1.03	c-C
Mississippi	Anoka, MN	missip1	49448	19	5.2	13	53	23	9	1.2	0.76	s-C
Isle	France	isle	6568	4	10	14	29	24	12	0.66	0.54	c-C
Seine	Choisy, France	choisy	30710	10	7.6	20	55	25	7	1.45	0.89	s-C
Green	Green, UT	gre	116124	42	1.4	283	4245	25	11	0.67	0.96	c-C
Iowa	Wappelo, IA	iow1	8468	18	8.5	109	269	26	9	0.55	0.37	c-C
Rio Grande	Otowi Bridge, NM	rio	37025	33	1.1	132	2520	26	12	0.33	0.63	c-C
Minnesota	Mankato, OH	min	38579	27	3.4	82	506	27	15	0.35	0.39	c-C
Tennessee	Savannah, TN	ten2	85796	7	15.9	121	160	28	10	0.91	0.83	c-C
Scioto	Chillicothe, OH	sci	9982	7	10.7	106	208	28	12	0.9	0.73	c-C
Columbia	Vancouver, WA	clb	668000	6	9.16	25	57	30	6	1.7	1.6	c-C
Iowa	Wappelo, IA	iow2	32358	10	8.3	206	521	32	8	0.32	0.81	c-C
Green	Jensen, UT	gre1	76795	30	1.6	175	2297	34	11	1.12	0.92	c-C
Dordogne	à Pessac, France	dordogne	14925	4	26.5	58	46	35	8	1.46	1.08	c-C
Garonne	France	garonne	51500	4	11.2	38	71	35	8	1.83	0.92	c-C
Tradewater	Olney, KY	trd	660	6	14.5	50	72	35	18	0.68	0.36	c-C
Green	Mudfordville, KY	gre2	4331	11	20.2	266	277	36	14	0.84	0.64	c-C
Brazos	Richmond, TX	brz	116518	20	1.65	131	1667	36	15	0.97	1.19	c-C
Arkansas	Las Animas, CO	arkanim	37340	22	0.7	19	570	37	14	0.61	0.73	c-C
Little Black	Success, MO	lbl	502	5	13.8	30	46	38	20	-0.1	0.04	s-S
East Fork	Seymour, IN	eastfork	6061	13	12.8	174	285	39	16	0.58	0.5	c-C
Tennessee	Chattanooga, TN	ten3	55402	7	15.9	257	339	39	11	1.21	1.07	c-C
Cuyahoga	Independence, OH	cuy	1834	22	15.1	301	419	41	10	0.97	0.91	c-C
Cuyahoga	Old Portage, OH	cuy2	1046	8	14.3	81	119	41	10	0.97	0.83	c-C
Sandusky	Fremont, OH	san	3245	12	11.1	214	405	42	19	0.81	0.62	c-C
San Juan	Shiprock, NM	sanjuan	33400	31	1.7	552	6819	43	11	0.35	0.59	c-C
Garonne	France	garonne	51500	6	11.6	38	69	44	10	1.83	0.9	c-C
Maumee	Waterville, OH	mau	16395	20	9.4	167	373	44	19	0.84	0.44	s-C
Dan	Paces, VA	dan	6602	12	13.5	240	373	47	15	1.16	0.81	c-C
Arkansas	Arkansas City, KS	arkan	113180	13	0.5	51	2142	47	24	0.91	0.88	c-C
Raisan	Monroe, MI	rai	2698	5	0.2	63	6615	49	15	0.92	0.42	s-C
Grand	Painsville, OH	grd	1773	12	17.5	215	258	49	17	0.99	0.45	s-C
Grand	Painsville, OH	grd2	1773	11	17.1	80	98	49	17	0.99	0.37	s-C
Gave	France	gave	5430	4	31.9	75	49	49	9	2.07	1.18	c-C
Iserre	France	isere	5700	5	29.8	471	332	51	6	2.81	1.74	c-C
Delaware	Trenton, NJ	del	17553	14	20.5	99	101	55	10	1.38	0.88	c-C
Conococheague	Fairview, MD	cono	1279	13	16.4	126	161	55	16	1.11	0.75	c-C
Fisher	Libby, MT	fis	2169	8	7.5	113	316	56	13	1.3	1.25	c-C
Juniata	Newport, PA	jun	8684	38	13.6	74	114	57	15	1.24	0.86	c-C
Muddy	Vaughn, MT	mud	730	12	4.5	277	1293	59	11	1.35	0.77	s-C
Mad	Arcata, CA	mad	1256	7	34.4	4725	2884	59	22	1.36	1	c-C
Coal	Alum Creek, AZ	coal	2162	4	16	552	725	61	20	1.29	0.98	c-C
Trinity	Hoopla, CA	tri	7386	9	19.2	1264	1383	62	21	1.47	1.42	c-C
Suislaw	Mapleton, OR	sui	1522	7	46	325	148	62	17	1.13	0.6	c-C
Feather	Gridley, CA	fea	9517	24	22.1	15	14	63	18	0.85	0.28	c-C
Pembina	Walhalla, ND	pern	8674	13	1	122	2562	64	32	0.93	0.47	s-C
Eel	Fort Seward, CA	eelfort	5455	9	27.5	4135	3158	65	28	1.44	1.02	c-C
Rappahannock	Remington, VA	rappah	1605	41	11.7	140	251	66	19	1.35	0.68	c-C
Brandywine	Chadds Ford, PA	brandy	813	33	17.3	165	200	69	14	1.66	0.81	s-C
Redwood	Orick, CA	red	717	14	39.1	3557	1910	69	22	1.19	1.03	c-C
Pecos	Santa Rosa, NM	pecos	6861	22	0.4	151	7928	70	33	1.42	1.18	c-C
Klamath	Orleans, CA	kla	21943	12	12.1	392	680	71	16	1.82	1.36	c-C
Gila	Solomon, AZ	gila						71	26	1	1.29	c-C
Upper Iowa	Dorchester, IA	uiow	1993	5	6.3	450	1500	75	18	1.57	1.13	s-C
Eel	Scotia, CA	eelscot	8060	20	28.3	7309	5424	76	30	1.52	1.13	c-C
Creek	Creek, CA	creek						81	23	0.89	0.44	s-C
San Pedro	Charleston, AR	spedro	3195	11	0.4	342	17955	82	48	1.37	1.44	s-C
Arkansas	Kinsley, KS	arkin	85604	15	0.05	3	1260	83	27	0.87	0.86	c-C
Paria	Leesferry, AR	paria	3651	27	0.2	1597	167685	83	34	2.1	2	c-C
Sanluisrey	Oceanside, CA	sluis	1443	9	0.6	34.2	1197	96	44	0.55	0.36	s-C
Santa Clara	Los Angeles, CA	sclara	1618	9	1.2	1667	29173	98	61	0.88	1.07	c-C

## 2.1.2 Stations nutriments dissous et totaux

La base de donnée sur les nutriments regroupe les nutriments dissous (ammonium, nitrate et phosphore dissous) et les nutriments totaux (phosphore total, azote Kjeldahl). Les données de concentrations et de débits journaliers sont issues de stations localisées sur les tributaires du Lac Erié. Pour l'ammonium et les nitrates, nous disposons également des données pour trois rivières françaises : l'Oise à Méry, la Marne à Neuilly et la Seine à Choisy, qui sont localisées sur la figure 2-2. Seule la station Vermilion aux USA a été éliminée de la base de données car les données n'étaient pas exploitables.

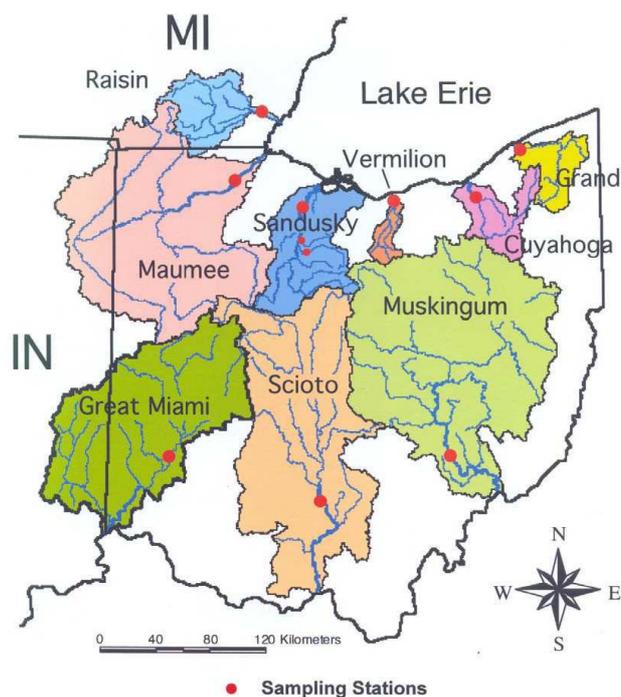


Figure 2-2 : Localisation des stations nutriments issues de la base de données des tributaires du Lac Erié

La gamme de bassin versant est moins étendue que pour les MES, allant de 1 800 à 31 000 km<sup>2</sup>. Les caractéristiques de débit et charge spécifiques ainsi que l'occupation du sol sont indiquées par station dans le tableau 2-2.

Les tributaires dans la partie ouest du lac Erie sont caractérisés par un régime hydrologique influencé par la fonte des neiges et un couvert forestier important, avec des concentrations plus faibles en phosphate et nitrate, mais plus important en azote Kjeldahl. Les rivières Grand et Cuyahoga sont impactées par les activités urbaines et industrielles (Richards, 2002).

Les autres rivières tributaires du lac Erie ainsi que les stations du bassin de la Seine sont plus agricoles, ce qui explique les concentrations plus importantes en nitrate (concentration moyenne inter-annuelle pondérée par les débits entre 13 et 18 mg/l NO<sub>3</sub>). Les caractéristiques des stations diffèrent selon les nutriments avec une variabilité des flux,  $M_{2\%}$ , variant de 7 à 24 % pour l'ammonium, 6 à 21% pour les nitrates, 13 à 76% pour le phosphore dissous et de 17 à 36% pour les nutriments totaux. Le comportement des concentrations en nutriments avec les débits est rarement linéaire et ainsi plusieurs types de relation C-Q sont représentés :

Nutriments dissous.

La base de données en nutriments dissous (20 stations) présente sept types de relation C-Q : **s-S** (2 stations), **s-C** (1 station), **d-D** (4 stations), **d-S** (4 stations), **c-D** (1 station), **c-S** (7 stations), **c-C** (1 stations)

Nutriments totaux.

Sur les 14 stations de la base de données en nutriments dissous, trois types de relation C-Q sont représentées : **s-C** (5 stations), **d-C** (6 stations), **c-C** (3 stations)

Tableau 2-2 : Caractéristiques générales des stations nutriments

Rivière	Stations	Var	BV (km²)	An	Q* (l/s/km²)	Y* (t/an/km²)	C* (mg/l)	Occupation du sol				M <sub>2%</sub> (%)	W <sub>2%</sub> (%)	b <sub>50sup</sub>	b	Typologie Relation C-Q
								Agri. (%)	Urb (%)	Forêt (%)	Autre (%)					
Oise	Mery, France	NH4	16972	10	7.2	0.05	0.1	71	6	22.5	0.5	7	8	-0.37	-0.2	d-D
Seine	Choisy, France	NH4	30710	10	7.6	0.02	0.1	58.7	4.9	35.6	0.8	14	7	-0.04	-0.04	d-S
Marne	Neuilly, France	NH4	12710	10	9.5	0.05	0.1	69.2	4.2	25.9	0.7	24	8	-0.31	-0.3	d-D
Oise	Mery, France	NO3	16972	10	7.2	4.84	14.1	71	6	22.5	0.5	6	8	-0.16	0.02	c-D
Seine	Choisy, France	NO3	30710	10	7.6	6.2	17.1	58.7	4.9	35.6	0.8	7	7	-0.03	0.14	c-S
Marne	Neuilly, France	NO3	12710	8	9.5	6.8	15.0	69.2	4.2	25.9	0.7	8	8	0.06	0.16	c-S
Cuyahoga	Independence, OH	NO3	1834	22	15.1	3.86	5.4	30.4	9.6	50.1	9.9	8	12	-0.42	-0.51	d-D
Muskingum	Mc Connelville, OH	NO3	19218	9	11.8	8.1	14.4	52	1.7	43.4	2.9	11	9	0.16	0.35	c-S
Scioto	Chillicothe, OH	NO3	9982	7	10.7	7.2	14.1	80.2	4.6	12.9	2.3	14	12	0.06	0.15	s-S
Great Miami	Miamisburg, OH	NO3	6954	6	12.2	7.5	12.9	82.1	4.7	10.3	2.9	17	14	0.06	0.13	s-S
Maumee	Watterville, OH	NO3	16395	21	9.5	8.26	18.3	89.9	1.2	7.3	1.6	17	16	0.08	0.8	c-S
Raisan	Monroe, MI	NO3	2699	21	8.4	5.29	13.2	79	2.3	14	4.7	18	14	0.32	0.78	c-S
Sandusky	Fremont, OH	NO3	3245	12	11.1	9.48	17.9	84.1	0.9	13	2	21	19	0.03	0.73	c-S
Scioto	Chillicothe, OH	PO4	9982	7	10.7	0.12	0.2	80.2	4.6	12.9	2.3	13	12	-0.02	-0.19	d-S
Great Miami	Miamisburg, OH	PO4	6954	6	12.2	0.17	0.3	82.1	4.7	10.3	2.9	18	14	0.02	-0.3	d-S
Maumee	Watterville, OH	PO4	16395	21	9.5	0.05	0.1	89.9	1.2	7.3	1.6	22	16	0.17	0.41	c-C
Raisan	Monroe, MI	PO4	2699	21	8.4	0.02	0.1	79	2.3	14	4.7	30	14	0.51	0.16	s-C
Sandusky	Fremont, OH	PO4	3245	12	11.1	0.04	0.1	84.1	0.9	13	2	31	19	0.24	0.6	c-C
Muskingum	Mc Connelville, OH	PO4	19218	9	11.8	0.02	0.0	52	1.7	43.4	2.9	54	9	-0.16	-0.2	d-D
Grand	Painesville, OH	PO4	1777	9	15	0.01	0.0	40	0.9	45.2	13.1	76	17	0.01	-0.2	d-S
Scioto	Chillicothe, OH	Ptot	9982	7	10.7	0.3	0.6	80.2	4.6	12.9	2.3	17	12	0.24	0	d-C
Muskingum	Mc Connelville, OH	Ptot	19218	9	11.8	0.19	0.3	52	1.7	43.4	2.9	20	9	0.53	0.2	s-C
Great Miami	Miamisburg, OH	Ptot	6954	6	12.2	0.41	0.7	82.1	4.7	10.3	2.9	24	14	0.37	-0.09	d-C
Raisan	Monroe, MI	Ptot	2699	21	8.4	0.14	0.4	79	2.3	14	4.7	30	14	0.67	0.16	d-C
Maumee	Watterville, OH	Ptot	16395	21	9.5	0.34	0.8	89.9	1.2	7.3	1.6	31	16	0.49	0.25	s-C
Sandusky	Fremont, OH	Ptot	3245	12	11.1	0.36	0.7	84.1	0.9	13	2	33	19	0.52	0.41	c-C
Grand	Painesville, OH	Ptot	1777	9	15	0.17	0.2	40	0.9	45.2	13.1	36	17	0.5	0.11	d-C
Muskingum	Mc Connelville, OH	TKN	19218	9	11.8	0.75	1.3	52	1.7	43.4	2.9	17	9	0.38	0.05	s-C
Scioto	Chillicothe, OH	TKN	9982	7	10.7	1.06	2.1	80.2	4.6	12.9	2.3	18	12	0.31	0.13	s-C
Raisan	Monroe, MI	TKN	2699	21	8.4	0.84	2.1	79	2.3	14	4.7	22	14	0.33	0.13	d-C
Great Miami	Miamisburg, OH	TKN	6954	6	12.2	1.31	2.3	82.1	4.7	10.3	2.9	22	14	0.42	0.07	d-C
Cuyahoga	Independence, OH	TKN	1834	22	15.1	1.51	2.1	30.4	9.6	50.1	9.9	22	12	0.23	0.12	c-C
Maumee	Watterville, OH	TKN	16395	21	9.5	1.45	3.2	89.9	1.2	7.3	1.6	26	16	0.31	0.08	s-C
Sandusky	Fremont, OH	TKN	3245	12	11.1	1.67	3.2	84.1	0.9	13	2	29	19	0.36	0.22	c-C

## 2.1.3 Stations sels dissous totaux

La conductivité électrique à 25°C est considérée ici comme un proxy de la concentration en sels dissous totaux comme cela a été démontré pour les rivières françaises (Meybeck, 1986) et est généralement utilisée en surveillance de la qualité des eaux (Chapman, 1996). Comme l'assemblage ionique ne varie pas beaucoup aux stations, la corrélation entre la conductivité à 25°C en  $\mu\text{S}/\text{cm}$  et la somme des ions majeurs ( $\Sigma = \text{Ca}^{2+} + \text{Mg}^{2+} + \text{Na}^+ + \text{K}^+ + \text{Cl}^- + \text{SO}_4^{2-} + \text{HCO}_3^-$ ) exprimé en meq/l est considérée linéaire et stable pour une station donnée. Quelques rivières sont fortement concentrées comme Dolores (Utah) car NaCl domine aux faibles débits,  $\text{Ca}^{2+} \text{SO}_4^{2-}$  au débit moyen et  $\text{Ca}^{2+} \text{HCO}_3^-$  aux forts débits. La relation conductivité-sels dissous totaux varie. Cependant, l'influence de l'assemblage ionique reste exceptionnelle et assez limitée devant les variations de la conductivité

avec le débit. Dans la suite de ce travail, nous allons utiliser l'annotation SDT pour la conductivité électrique proxy des sels dissous totaux.

La base de données est constituée de rivières américaines (Figure 2-3). Ces stations couvrent une gamme très large de conditions hydrologiques, climatiques, lithologiques, d'occupations du sol. Les caractéristiques de chaque station sont détaillées dans le tableau 2-5 où les stations sont classées par ordre croissant de l'indicateur de durée des flux de matière,  $M_{2\%}$ , croissant.

La base de données en sels dissous totaux est donc composée de 47 stations avec 4 à 27 ans de données couvrant des gammes de taille de bassins versant de 700 à 1 100 000 km<sup>2</sup> (médiane= 13 000 km<sup>2</sup>). La médiane des débits spécifiques est plus faible par rapport aux stations MES et nutriments, avec un minimum de 0.01 l/s/km<sup>2</sup> pour Canadian R. Texas. En effet nous avons souhaité une gamme de concentration très étendue aussi bien pour les matières solides en suspension que pour les sels dissous totaux. Les valeurs extrêmes de conductivité électrique dépassent les 10 000  $\mu\text{S}/\text{cm}$  pour les conditions arides ou semi-arides où il y a formation d'évaporite (Texas, Arizona et Colorado).

La gamme de variation du  $M_{2\%}$  est moins étendue que pour les autres paramètres avec des valeurs allant de 3 à 19%. La gamme de taille des bassins versants est quant à elle très étendue. Les différentes caractéristiques par stations sont présentées dans le tableau 2-3. Sur les 47 stations de la base de données sels dissous totaux, cinq types sont représentées avec une majorité de type **d-D** (34 stations) puis **s-D** (5 stations), **c-D** (5 stations), **s-S** (2 stations), **d-S** (1 station),



Figure 2-3 : Localisation des stations SDT issues de la base de données USGS

Tableau 2-3 : Caractéristiques générales des stations SDT

Rivière	Stations	Code	BV	An	Q*	Y*	C*	M <sub>2%</sub>	W <sub>2%</sub>	b <sub>50sup</sub>	b	Typologie
			(Km <sup>2</sup> )		(l/s/km <sup>2</sup> )	(t/an/km <sup>2</sup> )	(mg/l)	(%)	(%)			Relation C-Q
Ocm	Ocmulgee, GA	ocm	6964	12	8.2	84.2	216	3	3	-0.69	-0.51	d - D
Colorado	Lee ferry, AZ	colferry	289560	10	1.4	84.5	1268	4	4	-0.07	-0.12	d - D
Missouri	Garrison, ND	misgar	44415	10	1.3	71.2	1151	4	3	0.03	0.03	s - S
Columbia	Quincy, OR	clb	665084	9	9.3	108	244	4	4	-0.21	-0.17	s - D
Pecos	Pecos, TX	pecos	91074	4	1.9	433	4788	5	10	-0.41	0.06	s - D
Missouri	Nebraska city, NE	misneb	1061441	27	1	57	1198	5	7	-0.24	-0.09	d - D
Gunnison	Gd Junction, CO	gunnison	20525	10	3.8	213	1178	6	10	-0.52	-0.48	s - D
Colorado	Cisco, UT	colcisco	62392	21	3.3	219	1394	6	12	-0.53	-0.49	d - D
Snake	Snake, ID	snake	14891	19	13	275	444	6	7	-0.14	-0.21	d - D
Oise	Mery, France	oise	16972	8	7.7	372	1015	7	7	-0.19	-0.09	s - D
Green	Green, UT	green	116111	14	1.2	55.6	973	7	11	-0.4	-0.28	d - D
Marne	Neuilly, France	marne	12710	10	9.5	395	874	7	8	-0.05	-0.02	d - D
Delaware	Trenton, NJ	delaware	17553	11	18.1	235	273	7	11	-0.3	-0.29	d - D
Clackamas	Estacada, OR	clack	1737	4	37.6	147	82	8	10	-0.18	-0.31	d - D
N Santiam	Mehama, OR	nsantiam	1694	6	52.3	145	58	8	9	-0.2	-0.19	d - D
Peace	Zolfo springs, FL	peace	2138	5	5.7	155	571	9	14	-0.35	-0.33	d - D
Arkansas	Avondale, CO	arkavon	16380	20	1.6	89	1169	9	11	-0.26	-0.33	d - D
Arkansas	Granite, CO	arkgran	1105	13	9.6	87	190	9	12	-0.21	-0.28	d - D
Dolores	Cisco, UT	dolores	11857	8	1.9	156	1725	9	22	-0.64	-0.61	d - D
Sun	Vaughn, MT	sun	4787	7	3.7	179	1016	10	15	-0.3	-0.27	d - D
Duchesne	Duchesne, UT	duchesne	11000	14	1.5	109	1527	10	18	-0.41	-0.36	d - D
Sjoaq	San Joaquin, CA	sjoaq	25097	11	2.8	99	743	10	20	-0.52	-0.5	d - D
Arkansas	Ralston, OK	arkral	141003	9	1.1	99	1891	10	24	-0.41	-0.25	c - D
Ninnescah	KS	nines	1846	8	1.8	148	1727	10	21	-0.34	-0.1	c - D
San juan	Bluff, UT	sjuan	59544	6	1	62	1303	11	8	-0.24	-0.38	d - D
Brandywine	Chadds Ford	brandy	743	8	18.9	328	365	11	15	-0.19	-0.18	d - D
Seine	Choisy, France	choisy	30710	10	7.6	292	807	11	16	-0.21	0	c - D
Potomac	Wash, DC, MA	potom	29927	10	13.3	278	439	11	16	-0.21	-0.18	d - D
North Canadian	Britton, OK	ncan	34725	8	0.2	9.5	998	12	18	0.03	-0.02	c - D
North Canadian	Yukon, OK	ncanyukon	34129	8	0.2	19	1996	13	19	-0.1	-0.02	s - D
Jackson	Gath	jacksongath	1061	9	13.8	180	274	14	18	-0.16	-0.23	d - D
Shenandoah	Front royal, VA	shenan	4230	8	12	239	418	14	17	-0.17	-0.23	d - D
Virgin	Littlefield, AZ	littlefield	13177	10	0.6	102	3572	14	20	-0.39	-0.32	d - D
Arkansas	Las animas, CO	arkanim	37324	8	0.3	40	2801	16	25	-0.33	-0.32	d - D
Wichita	Charlie, TX	wichitachar	8903	5	0.8	189	4964	16	21	-0.31	-0.23	d - D
Jackson	Falling springs, VA	jacksonfal	1061	9	13.8	180	274	17	18	-0.04	-0.23	d - S
Nwchita	Truscott, TX	nwich	2426	12	0.6	386	13516	17	48	-0.48	-0.33	d - D
Nwchita	Paducah, TX	nwichpad	1398	12	0.5	540	22691	18	48	-0.45	-0.4	d - D
Wichita	Wichita falls, TX	wichita	8129	5	0.5	140	5883	18	28	-0.37	-0.3	d - D
Rappahannock	Remington, VA	rappah	1603	10	13.3	72	114	18	20	-0.04	-0.07	d - D
Canadian	Amarillo, TX	canamar	50341	6	0.01	6	12606	19	35	-0.41	-0.07	c - D
Medina	San Antonio, TX	medina	3410	12	2.4	133	1164	19	32	-0.24	-0.22	d - D
Bird	Catoosa, OK	bird	2646	6	6.3	146	487	19	24	-0.12	-0.13	d - D
Sheyenne	Lisbon, ND	shey	21203	7	0.3	16	1121	22	27	-0.12	-0.13	d - D
Meherrin	Emporia, VA	meh	1926	4	13.1	106	170	23	22	-0.15	-0.07	s - S
Knife	Hazen, ND	knife	5799	6	0.8	56	1471	24	48	-0.25	-0.21	d - D
Redwood	Burkburnett, TX	red	53253	8	0.9	315	7354	27	37	-0.3	-0.19	d - D

### 2.1.4 Stations du bassin de la Seine

Deux types de données sont disponibles sur le bassin de la Seine : (i) les données de concentrations journalières provenant du Syndicat des eaux d'Île de France (SEDIF) ; (ii) les données « qualité » mesurées de manière discrète à une fréquence mensuelle ou bimensuelle par l'Agence de l'Eau Seine Normandie (AESN).

Les données du SEDIF concernent trois stations, la Seine à Choisy, l'Oise à Méry et la Marne à Neuilly (figure 2-10) pour les MES, les nutriments et les SDT (cf. tableaux 2-1, 2-2, 2-3). Ces données sont exploitées dans la démarche générale de la thèse pour évaluer les performances des méthodes de calcul.

Sur l'ensemble des 279 stations de qualité et hydrométriques disponibles sur le bassin de la Seine, une première sélection a été réalisée en fonction du nombre de mesures de MES et de nitrates à chaque station. Ainsi pour les stations de qualité, ont été retenues celles disposant d'au moins 48 mesures sur l'intégralité de la période de suivi. Pour les stations hydrométriques, ont été sélectionnées celles qui disposaient d'au moins 350 mesures par an sur un minimum de 10 ans. Les associations entre stations de qualité et hydrométriques a ensuite été réalisées sur des critères de taille de bassins versants. La station hydrométrique est associée à la station de qualité la plus proche si la différence de taille du bassin versant entre les deux types de stations n'excède pas à  $\pm 30\%$ . Ainsi 279 stations de réseau de surveillance de la Seine ont été sélectionnées. Une deuxième sélection avec des bassins versants supérieurs à 500 km<sup>2</sup> a été réalisée et 64 stations ont ainsi été retenues pour l'ammonium (NH<sub>4</sub>), les nitrates (NO<sub>3</sub>), le phosphore total (P<sub>tot</sub>) et les matières en suspension (MES). Ces stations disposent de mesures depuis les années 1970 où les fréquences de suivis vont de bimensuelles (3 stations) à bi-trimestrielles (21 stations) (Figure 2-4). Nous ne nous intéresserons dans ce travail qu'à l'étude des flux de nitrate et de MES, une analyse critique des données de phosphore et d'ammonium est nécessaire avant de calculer les flux.

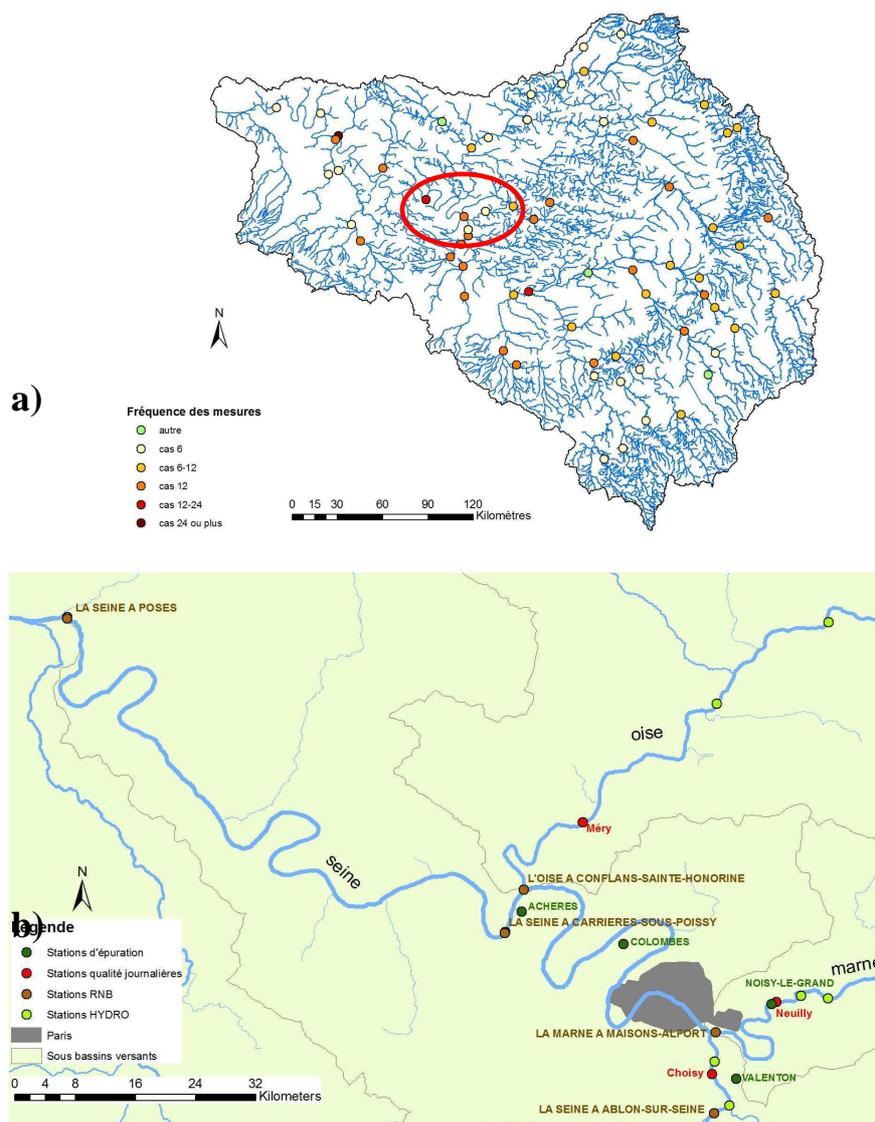


Figure 2-4 : Stations de qualité et hydrométriques utilisées sur le bassin de la Seine : a) ensemble du bassin avec les fréquences d'échantillonnage associées ; b) zoom sur les stations de qualité et hydrométriques utilisées pour l'analyse des tendances ainsi que l'emplacement des stations d'épuration.

La variabilité hydrologique ( $W_{2\%}$ ) et hydrochimique/sédimentologique ( $b_{50sup}$ ) est calculée pour chaque station à partir des débits en continu et des données discrètes des concentrations. Ces stations sont ensuite réparties dans la matrice de variabilité définie dans cette thèse. Le tableau 2-4 représente ainsi le nombre et la répartition par classe de  $W_{2\%}$  et  $b_{50sup}$  des stations pour chaque type de matériaux :

Tableau 2-4 : Répartition et effectif pour chaque classe -  $W_{2\%}$  /  $b_{50sup}$  - des stations sélectionnées

NO3	$b_{50sup} < -0.2$	$-0.2 < b_{50sup} < 0.2$	$0.2 < b_{50sup} < 0.8$	$0.8 < b_{50sup} < 1.4$	$b_{50sup} > 1.4$	Total
$W_{2\%} < 10$	2	25	8	0	0	35
$10 < W_{2\%} < 15$	1	24	1	0	0	26
$15 < W_{2\%} < 25$	0	3	0	0	0	3

MES	$b_{50sup} < -0.2$	$-0.2 < b_{50sup} < 0.2$	$0.2 < b_{50sup} < 0.8$	$0.8 < b_{50sup} < 1.4$	$b_{50sup} > 1.4$	Total
$W_{2\%} < 10$	0	5	8	14	8	35
$10 < W_{2\%} < 15$	0	1	7	14	4	26
$15 < W_{2\%} < 25$	0	0	0	3	0	3

<b>Total</b>	<b>3</b>	<b>58</b>	<b>24</b>	<b>31</b>	<b>12</b>	
--------------	----------	-----------	-----------	-----------	-----------	--

Toutes les gammes de variabilité hydrochimique/sédimentologique sont représentées. 45% des stations sont compris entre -0.2 et +0.2 ; 20% entre +0.2 et +0.8 ; 24% entre +0.8 et +1.4 et 11% représente les conditions extrêmes de dilution ( $< -0.2$ ) ou de concentration ( $> +1.4$ ). Trois gammes de variabilité hydrologique sont définies avec 55% des conditions de  $W_{2\%} < 10\%$  et 41% des conditions  $10\% < W_{2\%} < 15\%$ . Trois stations seulement possèdent des  $W_{2\%} > 15\%$ .

### 2.1.5 Répartition de la base de données en fonction des indicateurs de variabilité utilisés

A partir de la typologie des relations C-Q, élaborée par Meybeck et Moatar (2010), on peut séparer l'intégralité de la base de données en fonction du type de relation C-Q en ne tenant plus compte de la nature du constituant. Ainsi trois types se dégagent avec des effectifs plus importants (Tableau 2-5) : les constituants dilués (d-D) correspondant majoritairement aux sels dissous totaux ; et les constituants concentrés en période de hautes eaux (s-C et c-C) correspondant majoritairement aux MES et quelques nutriments totaux. Les nutriments dissous ont des comportements très variés, cependant nous avons des effectifs plus faibles compte tenu d'un nombre de stations réduit.

Tableau 2-5 : Répartition des effectifs par types de relation C-Q en fonction du débit médian

	$b_{50sup} < -0.2$ Dilué (D)	$-0.2 < b_{50sup} < 0.2$ Stable (S)	$b_{50sup} > 0.2$ Concentré (C)
$b_{50inf} < -0.2$ dilué (d)	38	5	6
$-0.2 < b_{50inf} < 0.2$ stable (s)	5	5	19
$b_{50inf} > 0.2$ concentré (c)	7	6	60

Dans ce travail, nous avons élaboré une matrice qui caractérise les variabilités hydrologique, hydrochimique et sédimentologique sur la base des indicateurs de variabilité  $W_{2\%}$  et  $b_{50sup}$ . Elle sera utilisée pour comparer les performances des méthodes. On observe des effectifs par classe de  $W_{2\%}$ - $b_{50sup}$  homogènes (de 5 à 15 stations par classes) à l'exception de deux conditions ( $b_{50sup}$  de -0.2 à 0.8 et  $W_{2\%} > 25$ ) qui seront notées comme non documentées (Tableau 2-6).

Tableau 2-6 : Répartition des effectifs par classes de  $W_{2\%}$ - $b_{50sup}$ 

	$b_{50sup} < -0.2$ Dilué	$-0.2 < b_{50sup} < 0.2$ Stable	$0.2 < b_{50sup} < 0.8$ Faiblement concentré	$0.8 < b_{50sup} < 1.4$ Concentré	$b_{50sup} > 1.4$ Fortement concentré
$W_{2\%} < 10$ Variabilité faible	9	12	6	11	7
$10 < W_{2\%} < 15$ Variabilité moyenne	9	5	15	10	1
$15 < W_{2\%} < 25$ Variabilité forte	10	14	8	13	3
$W_{2\%} > 25$ Variabilité très forte	7	0 Non documenté	0 Non documenté	6	4

## 2.2 Méthodologie

Deux méthodologies ont été utilisées dans ce travail en fonction de l'objet d'étude : i) évaluation des incertitudes des flux de polluants qui forme la partie centrale de cette thèse (Figure 2-5) et ii) estimation des incertitudes du quantile 90% des concentrations, métrique utilisée par les agences de l'eau pour évaluer la qualité des eaux. L'essentiel du travail étant basé sur les flux, seule la méthodologie sur les flux sera développée dans ce chapitre alors que celle sur les quantiles 90% des concentrations sera discutée au chapitre 7.

La méthodologie générale pour les flux est présentée à la figure 2-5. A partir d'une sélection de stations constituant une base de données conséquente, on génère des simulations de suivis discrets suivant différents intervalles d'échantillonnage par la technique de Monte-Carlo. Les flux annuels sont estimés à partir des simulations selon les différentes méthodes de calculs des flux. Ces flux annuels estimés sont comparés au flux de référence calculé à partir des données journalières. Les incertitudes déterminées à partir des erreurs sont ensuite analysées par intervalle d'échantillonnage, matériaux, méthodes et stations. Des indicateurs de variabilité hydrologique, hydrochimique et hydrosédimentaire sont estimés à partir des données journalières ou des simulations afin d'être reliés aux incertitudes. L'analyse de ces relations vise à construire un système expert d'utilisation optimale des méthodes de calculs.

Les étapes de la méthodologie sont décrites une à une dans les sections suivantes.

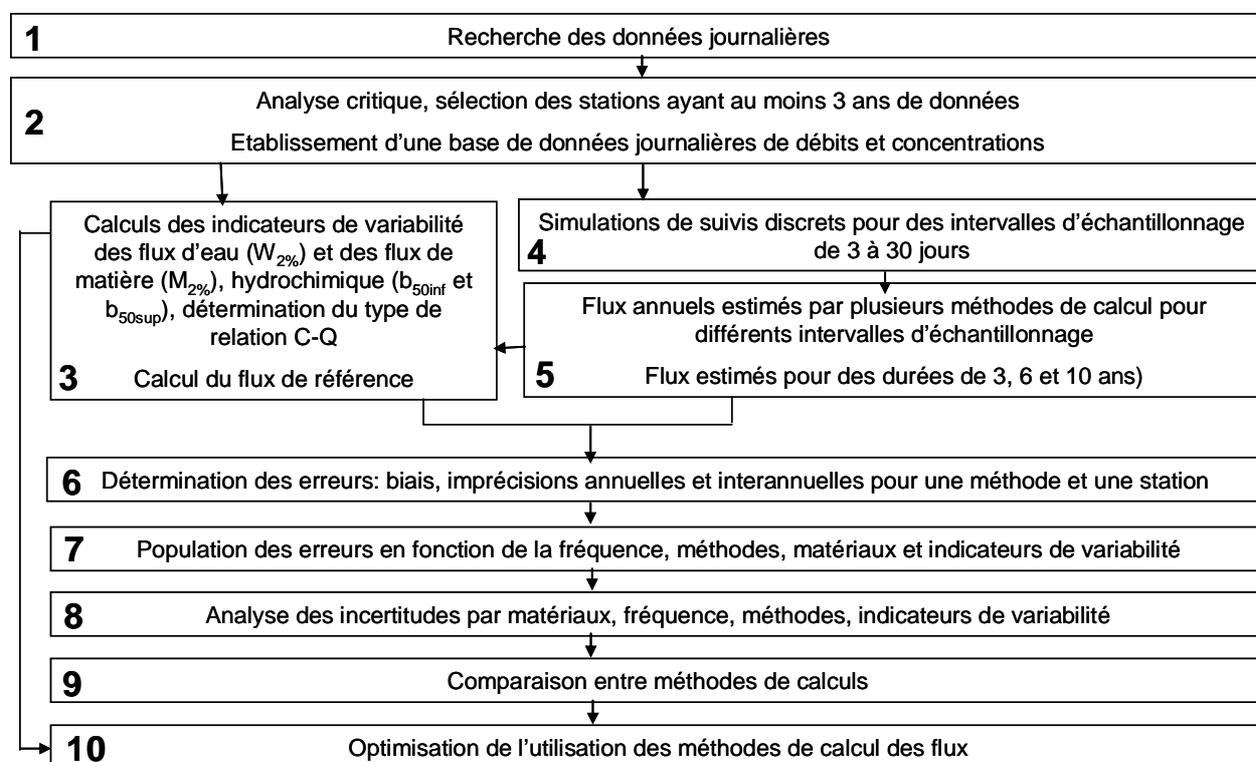


Figure 2-5 : Démarche scientifique générale pour l'étude des flux de polluants

L'ensemble des sections suivantes reprend les étapes de la méthodologie numérotées de 1 à 11.

### 2.2.1 Sélection des données (étapes 1 et 2)

La base de données couvre des matériaux très divers (matières en suspension, nutriments, sels dissous totaux) et des bassins versants contrastés. Pour couvrir la gamme de variabilité hydrologique interannuelle, nous avons privilégié les sites disposant de longues chroniques journalières (médiane 10 ans dans notre base de données). Nous avons aussi privilégié des réseaux de recherche établis sur du long terme qui permettent d'avoir des données homogènes (même type de prélèvement et mesure). Ces données proviennent pour les matières en suspension (MES) et les sels dissous totaux (SDT) de l'United States Geological Survey (USGS) (<http://co.water.usgs.gov/sediment/>), pour les nutriments du Ohio Tributary Monitoring Program (National Center for Water Quality Research, Heidelberg College, Ohio, USA) sur les tributaires du lac Erié et du SEDIF (Syndicat des Eaux d'Ile de France) pour les stations du bassin de la Seine. Les méthodes standardisées de collecte des données de l'USGS permettent de comparer les données entre les stations. Les concentrations moyennes journalières sont des valeurs pondérées par le temps (Porterfield, 1972). L'information détaillée sur les protocoles d'échantillonnage et la collecte des données est disponible dans Edwards et Glysson (1988) pour les MES.

### 2.2.2 Détermination des indicateurs de variabilité (étape 3)

Dans un premier temps, une série de descripteurs a été déterminée afin de caractériser la gamme de variation des concentrations, débits, flux au niveau des stations :

- $Q^*$ , débit spécifique (échelle annuelle, interannuelle) ( $m^3/s$ )
- $Q_{50}$ , débit spécifique médian, descripteur de la distribution des débits journaliers ( $m^3/s$ )
- $C^*$ , concentration moyenne pondérée par les débits (mg/l)
- $Y^*$ , charge spécifique des polluants en tonnes/an/km<sup>2</sup>
- $W_{2\%}$ , le % du flux d'eau s'écoulant en 2% du temps sur la période d'étude par débits classés décroissants. Il s'agit d'un descripteur de variabilité hydrologique pendant la période des hautes eaux. Il est ainsi déterminé pour l'ensemble des années ( $W_{2\%}$  interannuel) et par année ( $w_{2\%}$  annuel) sur la base des débits journaliers.
- $M_{2\%}$ , le % du flux de matière s'écoulant en 2% du temps sur la période d'étude par flux classés décroissants. Il s'agit d'un descripteur de variabilité des flux de matière. Il est ainsi déterminé pour l'ensemble des années ( $M_{2\%}$  interannuel) et par année ( $m_{2\%}$  annuel) sur la base des données journalières de concentrations et de débit. Des valeurs fortes de ces deux descripteurs ( $W_{2\%}$  et  $M_{2\%}$ ) vont ainsi définir une variabilité importante des flux de matière ou d'eau pour les rivières.
- $b$ , pente de la relation  $\log C - \log Q$ , pour l'intégralité des données

- $b_{50sup}$ , exposant de la régression Log C- Log Q, pour les valeurs supérieures au débit médian  $Q_{50}$ . Il peut être déterminé à partir des chroniques de surveillance simulées et également à partir des données journalières.
- $b_{50inf}$ , exposant de la régression Log C- Log Q, pour les valeurs inférieures au débit médian  $Q_{50}$ . Ces deux descripteurs ( $b_{50inf}$  et  $b_{50sup}$ ) vont servir à définir les types de relation C-Q tels que décrits au tableau 1-10 (chapitre 1).

### 2.3 Simulation des suivis discrets pour des intervalles d'échantillonnage de 3 à 30 jours (étape 4)

Afin de reproduire des suivis de surveillance à différents intervalles d'échantillonnage, des simulations de suivis discrets avec des intervalles de prélèvement de 1 à 30 jours ont été générées sur la base des données journalières par la technique de Monte-Carlo. Cette technique permet de créer aléatoirement un suivi discret possible aux intervalles souhaités. Par exemple, pour un intervalle mensuel, une date est tirée au sort à partir du 1<sup>er</sup> mois de la base de données puis ensuite un intervalle de moyenne 30 jours et d'écart type 4 jours (Figure 2-6).

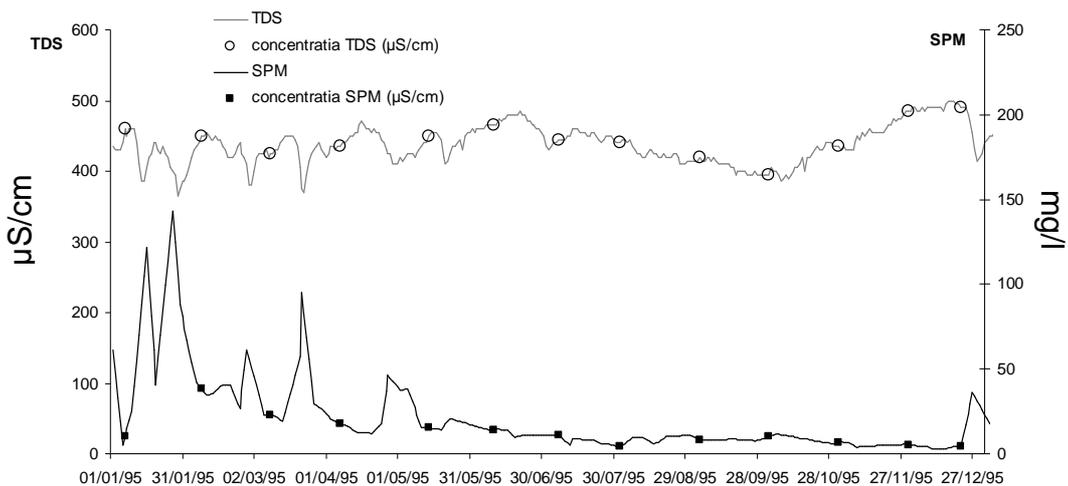


Figure 2-6 : Exemple d'un échantillonnage discret à un intervalle d'échantillonnage mensuel sur une année

Afin de traiter statistiquement les données, 50 simulations de suivis discrets sont générées pour un intervalle et une station. Le choix de 50 simulations ou « tirages » a été fait suite aux travaux de Moatar et Meybeck (2007) (Figure 2-7).

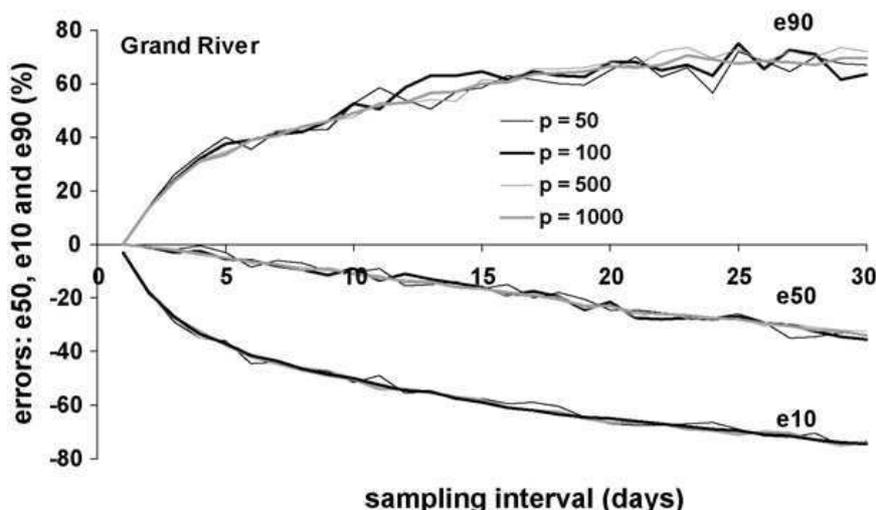


Figure 2-7: Evolution des quantiles des erreurs sur les flux annuels de MES (e10, e50, e90, valeurs interannuelles) pour la rivière Grand en fonction de l'intervalle d'échantillonnage (d, jours) et le nombre de tirages au sort ( $p = 50, 100, 500, 1000$ ) (Moatar et Meybeck, 2007).

Ces auteurs ont montré en effet que les quantiles 10, 50 et 90% des erreurs restent (figure 2-7) similaires pour des intervalles d'échantillonnage de 3 à 30 jours lorsque le nombre de simulations varie de 50 à 1000. Ce choix de 50 tirages est donc le compromis entre un résultat fiable et des temps de calcul raisonnable permettant l'utilisation de plusieurs méthodes à plusieurs stations.

### 2.3.1 Description des méthodes de calculs (étape 5)

Neuf méthodes de calculs des flux ont été retenues regroupant des méthodes fréquemment utilisées (méthodes de régression et de moyennes) comme des méthodes plus récentes (basée sur la distribution des concentrations). Elles sont classées ici en cinq catégories : i) les méthodes basées sur les moyennes des concentrations et des débits ; ii) les méthodes basées sur les régressions log C-log Q ; iii) les méthodes basées sur les régressions log C-log Q segmentées au débit médian ; iv) les méthodes basées sur les séparations d'hydrogramme ; v) les méthodes basées sur la distribution statistique des concentrations.

#### 2.3.1.1 Méthodes basées sur les moyennes (DWC et DWC\*)

La méthode du produit de la concentration moyenne pondérée par les débits et le débit moyen est utilisée dans cette thèse sous le nom de méthode **DWC** (pour *discharge weighted mean concentration method*). Elle sera également testé avec la correction de Beale (1962) notée **DWC\***. La méthode DWC est donc basée sur le calcul de la concentration moyenne pondérée par les débits qui est ensuite multiplié par le débit moyen :

$$Flux\ annuel = K \frac{\sum_{i=1}^n C_i Q_i}{\sum_{i=1}^n Q_i} \bar{Q} = K \times C \times \bar{Q}$$

avec :  $C_i$  = concentration instantané au moment du prélèvement  $i$  (mg/l),  $Q_i$  = débit instantané au moment du prélèvement  $i$  ( $m^3/s$ ) ;  $\bar{Q}$  = module annuel ( $m^3/s$ ), déterminé à l'aide de la chronique journalière des débits ;  $C^*$  = concentration moyenne pondérée par les débits (mg/l) ;  $n$  = nombre de prélèvements sur une année ;  $K$  = facteur de conversion du flux pour un résultat en tonnes/an (pour  $C_i$  en mg/l et  $Q_i$  en  $m^3/s$  ;  $K= 31.536$ )

Cette méthode a été utilisée entre autres par Verhoff et al., (1980) et Walling et Webbs (1981)

La méthode a été ensuite corrigée par Beale (1962), pour réduire le biais lié au manque de données sur la gamme complète de débit et de concentration, et se définit par :

$$Flux\ annuel = K \frac{\sum_{i=1}^n C_i Q_i}{\sum_{i=1}^n Q_i} \bar{Q} \times CF$$

Avec :

$$CF = \frac{\left[ 1 + \frac{1}{n} \left( \frac{Slq}{lq} \right) \right]}{\left[ 1 + \frac{1}{n} \left( \frac{Sq^2}{q^2} \right) \right]}$$

$$Slq = \frac{1}{n-1} \left[ \sum_{i=1}^n Q_i^2 C_i - (n \bar{q} \bar{l}) \right]$$

$$Sq^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n Q_i^2 - (n \bar{q}^2) \right]$$

$CF$  = facteur de correction de Beale (Littlewood, 1995) ;  $\bar{l}$  = flux moyen journalier pour les jours où les concentrations sont connues ;  $\bar{q}$  = débit moyen journalier pour les jours où les concentrations sont connues (en  $m^3/s$ )

### 2.3.1.2 Méthodes basées sur les régressions C-Q (IRC et IRC\*)

Les régressions sont effectuées sur l'intégralité des valeurs C-Q issues de la simulation d'échantillonnage (soit pour un intervalle d'échantillonnage mensuel sur 10 ans : 120 valeurs). Cette méthode est notée ici **IRC** (pour *Integral Rating Curve*). La méthode IRC consiste donc en une reconstitution des concentrations manquantes par une régression de la forme :

$$\log \hat{C} = a + b \log Q \text{ avec } a = \log \alpha \text{ et } b = \beta$$

Les flux annuels sont ensuite calculés en sommant le produit des débits et des concentrations estimées (pour les jours manquants) et des concentrations mesurées (pour les jours échantillonnés) :

$$\text{Flux annuel} = K \sum_{i=1}^{365-n} \hat{C}_i \times Qi + K \sum_{i=1}^n C_i \times Qi$$

avec  $\hat{C}_i$  = Concentration estimée d'après la régression (mg/l) ; a et b = paramètre de la régression.

Pour corriger la sous estimation due à la transformation logarithmique de la relation C-Q, Ferguson (1986) développe un facteur de correction qui est appliqué à la méthode IRC. Cette méthode corrigée est par la suite appelée la méthode **IRC\***. La correction de Ferguson se définit par :

$$\hat{C}_c = \alpha Q^\beta \times CF$$

$$CF = \exp(2.651 \sigma^2)$$

$\sigma$  correspond à l'écart type de l'estimation des concentrations

### 2.3.1.3 Méthodes basées sur les régressions segmentées (SRC<sub>50</sub> et SRC<sub>50</sub>\*)

Les relations log C- log Q n'étant pas toujours linéaires, on se propose de segmenter le jeu de données en fonction d'un critère Q<sub>50</sub>, correspondant au débit médian (Figure 2-8). Cette méthode est appelée par la suite **SRC<sub>50</sub>** (*Segmented Rating Curve*). La pente de la régression est calculée pour les valeurs en dessous ( $b_{50inf}$ ) et au dessus ( $b_{50sup}$ ) du débit médian, Q<sub>50</sub>.

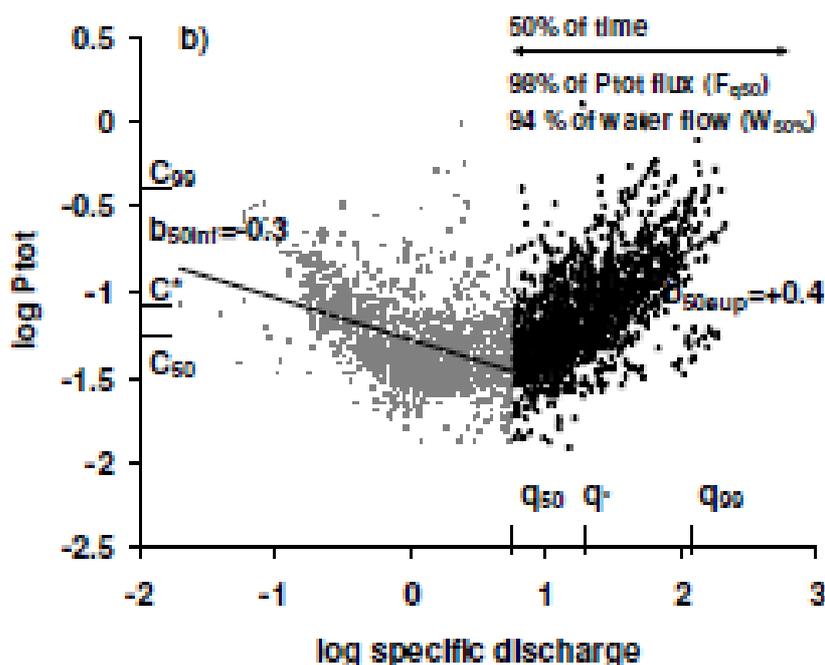


Figure 2-8: Régressions linéaires pour les valeurs en dessous et au dessus du débit médian avec le pourcentage du flux d'eau et de matière pour 50% du temps (d'après Meybeck et Moatar, 2010)

On estime donc les concentrations en fonction du débit en utilisant les régressions appropriées :

$$\log \hat{C} = a_{50inf} + b_{50inf} \log Q \text{ si } Q < Q_{50}$$

$$\log \hat{C} = a_{50sup} + b_{50sup} \log Q \text{ si } Q > Q_{50}$$

Les flux annuels sont ensuite calculés en sommant le produit des débits et des concentrations estimées (pour les jours manquants) et des concentrations mesurées (pour les jours échantillonnés). Comme pour la section 2.1.4.2, le facteur de correction de Ferguson est appliqué aux deux régressions effectuées. Cette méthode sera définie par la suite sous le nom de méthode **SRC50\***

### 2.3.1.4 Méthode basée sur la séparation d'hydrogramme (HS et HSM)

Une régression linéaire multiple est utilisée à partir des proportions des écoulements de base et de ruissellement au moment des prélèvements ( $\frac{Q_b}{Q}$  et  $\frac{Q_r}{Q}$  respectivement), et des concentrations mesurées. Cette méthode fait l'hypothèse d'une concentration constante dans les composantes de débit et sera dans la suite de la thèse nommée méthode **HS** (*Hydrograph Separation*). On a donc une équation de la forme :

$$\hat{C} = C_b \frac{Q_b}{Q} + C_r \frac{Q_r}{Q} + \varepsilon$$

avec  $\hat{C}$  = Concentration estimée (en mg/l) ;  $Q$  = Débit au jour du prélèvement (en m<sup>3</sup>/s) ;  $Q_b$  et  $Q_r$  = débit de base et de ruissellement (en m<sup>3</sup>/s) ;  $C_b$  et  $C_r$  = paramètres de la régression correspondant à la concentration dans le débit de base et dans le ruissellement. Les flux annuels sont ensuite calculés en sommant le produit des concentrations estimées et des débits journaliers.

D'après les travaux de Bustillo (2005) sur la Garonne, une nouvelle méthode de calcul des flux basée sur la séparation d'hydrogramme a été testée. Sur la base de la méthode HS, un module quadratique dépendant uniquement du ruissellement est ajouté à la régression multiple. Cette méthode est appelée par la suite la méthode **HSM** (*Hydrograph Separation Modified*). Ce module suggère une variation linéaire des concentrations avec le débit de ruissellement qui se traduit par :

$$C_r = C_{r1} + C_{r2} Q_r$$

Si l'on remplace dans l'équation de la méthode HS cela donne :

$$\hat{C} = C_b \frac{Q_b}{Q_i} + C_{r1} \frac{Q_r}{Q_i} + C_{r2} \frac{(Q_r^2)}{Q_i} + \varepsilon$$

Gonzales et al. (2009), dans une étude comparative des méthodes de séparation d'hydrogramme, ont montré que les meilleures méthodes sont les méthodes de *Rating Curve/enveloppe* (Sellinger, 1996 ; Kliner et Knezek, 1974) et les méthodes de filtrage autorécurssif d'Eckhardt (2005). Cette dernière a donc été employée dans ce travail pour séparer les écoulements et déterminer les deux composantes, ruissellement et débit de base, du débit total.

La méthode complète de séparation d'hydrogramme est détaillée dans le travail d'Eckhardt (2005) et se base sur les données journalières de débits pour l'année d'étude afin de distinguer la chronique des écoulements de base ( $Q_b$ ) et celle de ruissellement ( $Q_r$ ). Pour cela, on utilise la formule développée par Eckhardt (2005) :

$$Q_b(t) = \frac{(1 - BFI_{\max}) a Q_b(t-1) + (1 - a) BFI_{\max} Q(t)}{1 - a BFI_{\max}}$$

$$\text{avec } Q_r(t) = Q(t) - Q_b(t)$$

Où  $Q_b(t)$  = débit de base au temps  $t$  ;  $Q_b(t-1)$  = débit de base au temps  $t-1$  ;  $Q(t)$  = débit total au temps  $t$  ;  $BFI_{\max}$  = indice du débit de base (Indice définissant la proportion du débit de base sur le débit total) ;  $a$  = coefficient de tarissement correspondant à la vidange des nappes.

Une fois les concentrations reconstituées, le flux annuel est calculé comme précédemment en sommant le produit des concentrations estimées et des débits journaliers.

### 2.3.1.5 Méthode basée sur la distribution statistique des concentrations

Cette méthode proposée par Mailhot et al. (2008), appelée par la suite méthode **DC** (Distribution des Concentrations), considère la concentration comme une variable aléatoire qui peut être représenté par une distribution statistique dont les paramètres sont fonction des débits et du jour calendaire. Le modèle est donc défini par une loi de distribution (log-normal et Gamma dans la présente étude) et un modèle de moyenne décrivant la relation fonctionnelle entre les paramètres de la distribution et les covariables et un modèle de coefficient de variation des concentrations. Cette méthode a été évaluée au cours d'un stage de quatre mois au sein de l'INRS-ETE. Pour des raisons pratiques, la méthode a été simplifiée par rapport à la méthode originale de Mailhot et al. (2008) afin de pouvoir être comparée aux autres méthodes. Il est à noter que, dans la méthode originale, la méthode ne donne pas une valeur de flux mais une distribution de ce flux annuel. Par ailleurs la dépendance au jour julien n'a pu être testée pour des raisons pratiques mais également parce que l'utilisation des débits comme seule covariable a donné des résultats satisfaisants. Il faut également signaler que la méthode originale a été testée sur une rivière à forte composante nivale d'où l'intérêt de la dépendance aux jours juliens (Mailhot et al., 2008). La méthode se décompose en trois étapes.

#### 1- Sélection de la loi de probabilité.

Deux distributions statistiques ont été testées, à savoir la loi Log-Normal et la loi Gamma. Les fonctions de densité de probabilité des lois sont données respectivement par les expressions suivantes :

$$f_{LN}(s|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma s} e^{-\frac{(\ln s - \mu)^2}{2\sigma^2}}$$

$$f_G(s|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\beta s}$$

$\mu$ ,  $\sigma$ ,  $\alpha$  et  $\beta$  sont les paramètres des distributions définis.

Cette dépendance est définie à travers un modèle multiplicatif composé d'un modèle de moyenne et d'un modèle de coefficient de variation.

### 2- Modèles multiplicatifs : Evaluation des paramètres (moyenne, CV)

28 modèles par simulation sont testés : 2 distributions x 7 modèles de moyenne x 2 modèles de coefficient de variation (tableau 2- 7)

Tableau 2-7 : Description des distributions des modèles de moyennes et de coefficient de variation utilisés dans la méthode DC

Distribution statistique	Moyenne	Variance	
Log-normal	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$	
Gamma	$\alpha / \beta$	$\alpha / \beta^2$	
Modèle de moyenne ( $g1(Q)$ )	Formules $E(C Q) = a_0 \cdot g1(Q)$		Nb param
Constant	$a$		1
Linéaire	$a Q + b$		2
Quadratique	$a Q^2 + b Q + c$		3
Loi de puissance	$a Q^b + c$		3
Exponentielle de loi de puissance	$\exp(a Q^b + c)$		3
Gaussienne	$a (1 - \exp(-3 Q^2 / b^2))$		2
Gaussienne linéaire	$a (1 + b Q - \exp(-3 Q^2 / c^2))$		3
Modèle de CV ( $g2(Q)$ )	Formules $CV(C Q) = \frac{\sqrt{\text{var}(C Q)}}{E(C Q)}$		
	$CV(C Q) = \lambda \cdot g2(Q)$		
Constant	$k$		
linéaire	$kQ + b$		

$a$ ,  $b$ ,  $c$ ,  $k$  sont les paramètres des modèles à estimer.

Les paramètres de la moyenne et du coefficient de variation ( $a$ ,  $b$ ,  $c$ ,  $k$ ,  $a_0$  et  $\lambda$ ) sont estimés à partir de la simulation d'échantillonnage.

### 3- Sélection du meilleur modèle

Le « meilleur » des 28 modèles est sélectionné en déterminant le maximum de vraisemblance. On utilise le critère AIC (Akaike Information Criterion).

$$AIC = -2 \ln L(\theta) + 2k$$

Avec  $L(\theta)$  correspondant à la log vraisemblance et  $k$  le nombre de paramètres.

En plus du AIC, un autre critère a été ajouté par rapport à la méthode originale pour éliminer les modèles aberrants. En effet la méthode fait l'hypothèse que chaque valeur de concentration possède le même poids, les modèles sont donc fortement influencés par les bas débits, mais ce sont les forts débits qui déterminent l'essentiel des flux. On extrait donc d'une simulation d'échantillonnage donnée les valeurs maximales de concentration et de débit ( $C_m$  et  $Q_m$  respectivement) ainsi que l'écart type des concentrations,  $\sigma_c$ . Si la concentration estimée par le modèle évalué sur l'ensemble des valeurs pour le débit maximale des données respecte la condition suivante alors le modèle est retenu :

$$|C(Q_m) - C_m| < \sigma_c$$

Le modèle sélectionné doit donc répondre au critère précédent et avoir le AIC le plus petit.

#### 4- Application

Afin de pouvoir être comparé aux autres méthodes, seule la moyenne de la loi de distribution est utilisée pour reconstituer les concentrations à partir des débits journaliers.

Si l'on prend l'exemple d'une loi de distribution log-normal, les concentrations sont reconstituées par :

$$\hat{C} = \left( e^{\mu + \sigma^2 / 2} \right)$$

$$\text{avec } \mu = \ln \left( \frac{a_0 \cdot g1(Q)}{\sqrt{\ln(\lambda^2 \cdot g2(Q) + 1)}} \right)$$

$$\text{et } \sigma = \sqrt{\ln(\lambda^2 \cdot g2(Q) + 1)}$$

$a_0$  et  $\lambda$  sont constants,  $g1$  et  $g2$  correspondent au modèle de moyenne et au modèle de CV respectivement (tableau2- 1).

Le flux annuel est ensuite calculé en sommant le produit des concentrations estimées par les débits journaliers.

Toutes les méthodes décrites dans cette section seront reprises dans les chapitres suivants et nommées par leurs acronymes résumés au tableau 2-8.

Tableau 2-8 : Récapitulatif des méthodes testées et acronymes associés

Descriptions	Noms
Méthode des concentrations moyennes pondérée	<b>DWC</b>
Méthode des concentrations moyennes pondérée avec correction de Beale	<b>DWC*</b>
Méthode de régression intégrale	<b>IRC</b>
Méthode de régression intégrale avec correction de Ferguson	<b>IRC*</b>
Méthode de régression segmentée	<b>SRC<sub>50</sub></b>
Méthode de régression segmentée avec correction de Ferguson	<b>SRC50*</b>
Méthode de séparation d'hydrogramme	<b>HS</b>
Méthode de séparation d'hydrogramme modifié	<b>HSM</b>
Méthode de distribution statistique des concentrations	<b>DC</b>

### 2.3.2 Détermination des incertitudes sur les flux (étapes 7 et 8)

A partir de la base de données journalières de concentrations et de débits, les flux annuels sont calculés en sommant les produits des concentrations et débit journaliers sur une année. Ce flux ainsi calculé sera désigné comme le flux de référence. A partir des 50 simulations d'échantillonnage effectuées, autant de flux annuels sont estimés en fonction de différentes méthodes de calculs. Les erreurs relatives sont donc déterminées entre le flux annuel de référence et le flux annuel estimé selon :

$$\text{Erreurs relatives}(\%) = \left( \frac{\text{Flux estimé} - \text{Flux référence}}{\text{Flux référence}} \right) * 100$$

Nous disposons donc par exemple pour une station de 10 ans de données journalières, 500 erreurs (50 erreurs annuelles \* 10ans). A partir de ces erreurs (Figure 2-9) sont déduites les incertitudes selon Trois critères que sont le biais, l'imprécision ainsi que l'erreur quadratique moyenne (RMSE pour Root Mean Square Error).

Pour ces trois critères, il existe deux échelles d'étude définies par :

- l'échelle interannuelle correspondant dans notre exemple aux 500 erreurs sur les flux annuels (50 valeurs annuelles x 10 ans)
- l'échelle annuelle correspondant aux 50 erreurs par an sur les flux annuels

Ainsi le biais annuel se définit comme la médiane des 50 erreurs annuelles et le biais interannuel comme la médiane des erreurs pour toutes les années, ce qui correspondrait dans notre exemple à la médiane des 500 erreurs. L'imprécision se définit comme la différence entre le quantile 90% et le quantile 10% des erreurs, soit pour une année donc sur 50 valeurs (imprécisions annuelles) soit pour toutes les années donc sur les 500 erreurs (imprécisions interannuelles). Le RMSE quant à lui combine le biais et l'imprécision, il est calculé de la manière suivante :

$$RMSE = \sqrt{(\text{biais})^2 + (\text{imprécisions})^2}$$

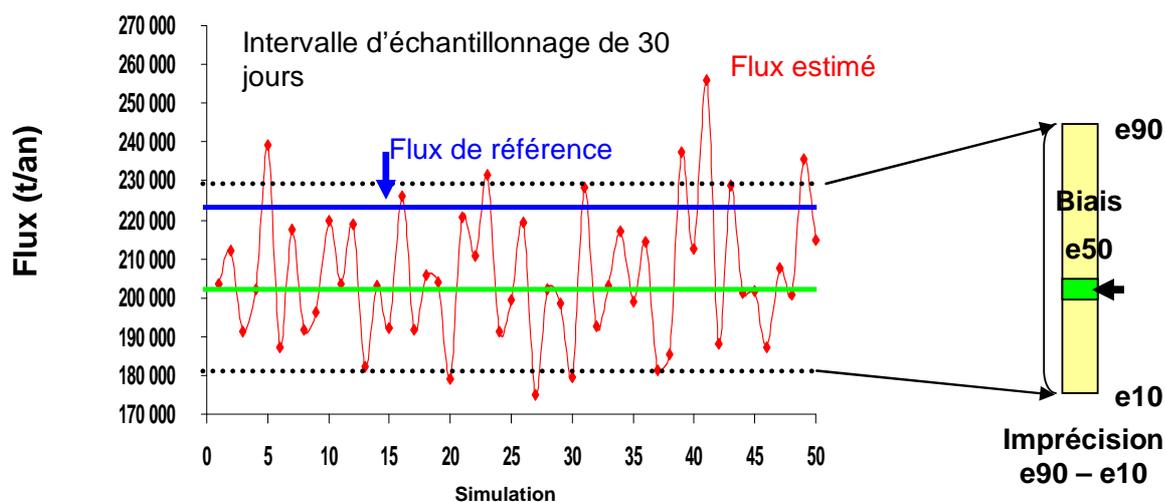


Figure 2-9: Détermination du biais et de l'imprécision à l'échelle interannuelle

Outre l'imprécision interannuelle, la moyenne des imprécisions annuelles est également calculée. L'imprécision interannuelle renseigne sur la dispersion des erreurs d'une méthode de calcul sur toutes les années étudiées sans prendre en compte la variabilité hydrologique d'une année sur l'autre, alors que la moyenne des imprécisions annuelle renseigne sur la dispersion des erreurs chaque année. Ceci donne deux informations différentes et complémentaires.

### 2.3.3 Analyse et comparaison des performances des méthodes de calcul (étapes 9 et 10)

Afin de comparer les performances des méthodes de calculs sur un grand nombre de stations, un emboîtement des différentes données est nécessaire. Cette démarche est définie à la figure 2-10. Les incertitudes sont donc représentées, dans un premier temps pour une station, un intervalle d'échantillonnage, une méthode et un type de matériaux (Figure 2-10a). Les erreurs sont calculées pour chaque année (An1, An 2...) et pour toutes les années confondues (Interan S1). Le niveau d'analyse supérieur est d'intégrer les incertitudes interannuelles pour toutes les stations pour un même paramètre de qualité de l'eau et pour une méthode (Figure 2-10b). Les erreurs interannuelles interstations sont calculées pour chaque méthode de calcul des flux et sont ainsi comparés sur cette base pour évaluer leurs performances (Figure 2-10c). Les boîtes à moustache représentées à la figure 2-10c intègrent donc les erreurs interannuelles pour toutes les stations pour un type de matériaux (MES, Nutriments...) sans tenir compte de leurs variabilités et leurs caractéristiques contrastées.

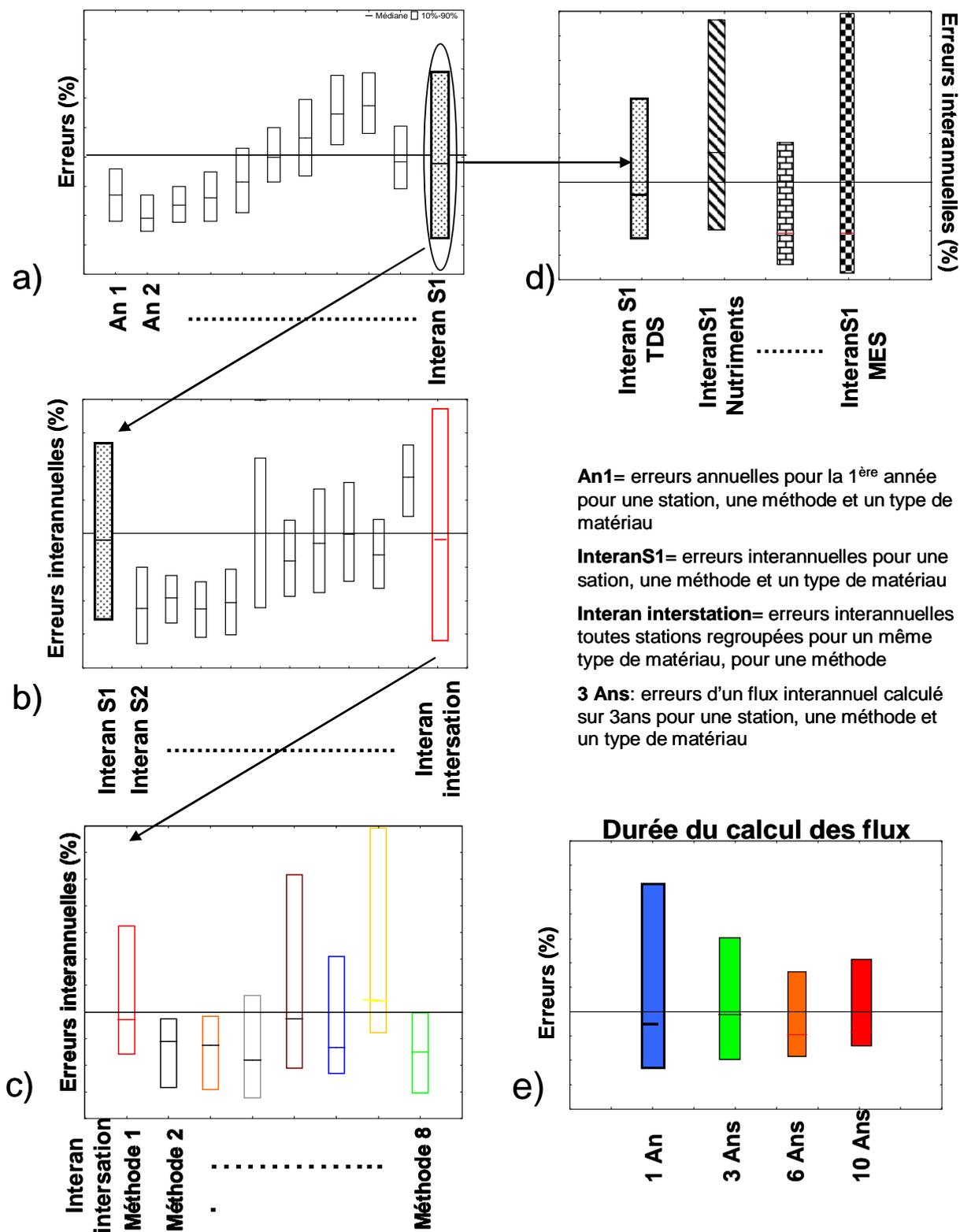


Figure 2-10: Evaluation des incertitudes : a) comparaison des erreurs annuelles pour une méthode, une station et un paramètre ; b) comparaison interstations des erreurs interannuelles pour une méthode et un paramètres ; c) comparaison des performances des méthodes à partir des erreurs interannuelles interstations ; d) comparaison des erreurs interannuelles pour une station, et une méthode pour des matériaux différents, e) erreurs sur les flux interannuels pour des durées de calcul de 1, 3, 6 et 10 ans.

Les erreurs interannuelles peuvent également être comparées pour une même station en fonction du paramètre de qualité de l'eau étudié (Figure 2-6d).

Toutes les étapes précédentes (2.1.2 à 2.1.5) ont ainsi été réalisées pour des simulations de surveillance de 3 à 30 jours afin de déterminer l'impact de la fréquence d'échantillonnage.

### 2.3.4 Influence de la durée du calcul des flux (étape 6 bis)

Toutes les étapes précédentes de la méthodologie se consacrent à l'étude des erreurs sur les flux annuels. Les erreurs ont donc été déterminées sur la même méthodologie mais avec des flux calculés sur 3, 6 et 10 ans. Ces calculs ont été réalisés afin de déterminer l'impact de la durée du calcul des flux sur les erreurs d'évaluation du flux de MES, NH<sub>4</sub>, Nitrate et SDT pour les rivières Oise, Marne et Seine.

### 2.3.5 Optimisation des méthodes de calcul des flux

Afin de déterminer les règles d'optimisation des flux pour l'ensemble des constituants dissous et particulaire, organique et inorganique, on s'abstrait de la nature du constituant pour ne retenir que sa variabilité statistique. Les stations ne sont ainsi plus regroupées par type de constituants mais en fonction de descripteurs  $W_{2\%}$  et  $b_{50sup}$ . Ainsi quatre classes de  $W_{2\%}$  et cinq classes de  $b_{50sup}$  ont été définies :

$W_{2\%}$  : <10% ; de 10% à 15% ; de 15% à 25% ; >25%

$b_{50sup}$  : < -0.2 : composant dilué, la plupart des SDT, ainsi que certains ortho phosphates et ammonium à l'aval de rejets d'eaux usées urbains ; de -0.2 à 0.2 : composant stable, certains SDT en région karstique ou des variations saisonnières comme pour les nitrates ; de 0.2 à 0.8 : composant faiblement concentré, le phosphore total et l'azote Kjeldahl ainsi que les MES pour les bassins de bas relief ; de 0.8 à 1.4 : composant concentré, MES dans les bassins à potentiel érosif modéré ; > 1.4 : composant très concentré, MES dans les bassins à fort potentiel érosif.

On définit ainsi une **matrice générale qui caractérise les variabilités des constituants et des débits**. Elle sera utilisée par la suite pour comparer les performances des méthodes et optimiser les fréquences d'échantillonnage. Une méthode est considérée comme acceptable lorsque les biais interannuels ( $e_{50}$ ) sont compris entre  $\pm 10\%$  et les imprécisions interannuelles sont dans la limite  $\pm 30\%$  ( $e_{90} - e_{10} < 60\%$ ).