



PRHYMO

Plateforme Pressions et Risques d'impacts HYdroMOrphologiques

Spécifications techniques

Aout 2024

Document associé à la version 062024.1.2

Présentation du document

Ce document fait partie d'un ensemble de documents qui accompagnent la diffusion des données de la version 062024.1.2 de PRHYMO :

- Rapport méthodologique : document principal qui décrit les principes de l'évaluation des pressions et risques d'altération hydromorphologiques, les méthodes utilisées et leurs évolutions par rapport aux précédents dispositifs SYRAH-CE et RHUM ;
- Rapport de comparaison : document qui compare les résultats des nouvelles évaluations avec les précédentes, issues des dispositifs SYRAH-CE et RHUM ;
- Spécifications techniques : document destiné aux administrateurs et gestionnaires de base de données pour prendre en main la banque de données, mettre à jour les données d'entrée puis relancer les calculs ;
- Scripts de mise en œuvre des algorithmes de calcul ;
- Document de présentation de la structure des arbres bayésiens : permet une visualisation rapide des modèles d'évaluation de chaque paramètre hydromorphologique ;
- Dictionnaire des attributs : décrit les attributs des données SIG fournies aux opérateurs, dans le cadre de la mise à jour des états des lieux DCE 2025 ;
- Sources de données : indique l'origine des données d'entrée utilisées.

Citation du document

Despres D., Grosprêtre L., 2024. Plateforme Pressions et Risques d'impacts HYdroMOrphologiques (PRHYMO) : spécifications techniques. Neogeo Technologies – Dynamique hydro, aout 2024.

Relecture - Validation

Karl Kreutzenberger (Office Français de la Biodiversité, OFB)

SOMMAIRE

SOMMAIRE	3
INTRODUCTION	4
1- SPECIFICATIONS A L'HEBERGEMENT ET A L'INSTALLATION	5
1.1. SPECIFICATIONS SERVEUR.....	5
1.2. SPECIFICATIONS LOGICIELLES.....	5
2- SPECIFICATIONS LIEES A L'ARCHITECTURE DE LA BANQUE DE DONNEES.....	5
2.1. STRUCTURE DES RESEAUX HYDROGRAPHIQUES	6
2.2. TABLES ISSUES DES USRA	8
2.3. LIENS AVEC LES BASSINS VERSANTS ET LES MASSES D'EAU	8
3- SPECIFICATIONS POUR L'EXECUTION DES CALCULS ET TRAITEMENTS	9
3.1. COLLECTE.....	9
3.2. TRAITEMENTS SQL	10
3.2.1. <i>Génération du réseau hydrographique en Hexagone</i>	10
3.2.2. <i>Calculs intermédiaires</i>	12
3.2.3. <i>Calculs des descripteurs</i>	13
3.3. TRAITEMENTS R	14
3.3.1. <i>Objectifs et principes</i>	14
3.3.2. <i>Architecture du compartiment bayésien</i>	15
3.3.3. <i>Chargement des CSV en banque de données</i>	19
3.3.4. <i>Lancement des calculs bayésiens</i>	19
3.4. RESULTATS	20
3.5. AUTOMATISATION DU LANCEMENT DES CALCULS.....	21
4- ANNEXE	23

INTRODUCTION

Ce document s'adresse en priorité à des administrateurs de (bases de) données, des géomaticiens ayant des compétences SIG avec PostGIS-PostgreSQL et des opérateurs manipulant le logiciel R.

L'objectif de ce document est de donner aux lecteurs la possibilité :

- De mettre à jour les données d'entrée nécessaires aux calculs ;
- De relancer la génération du réseau hydrographique numérique en Hexagone (en cas de mise à jour de la BD Topo par exemple ou de basculement vers la BD Topage) ;
- De recalculer les descripteurs à l'échelle de l'unité spatiale de recueil et d'analyse (USRA) ;
- De renouveler les calculs de risques d'altération à l'échelle de l'USRA et l'agrégation de ces risques à l'échelle du Tronçon géomorphologiquement homogène (TGH) et de la masse d'eau (ME).

Ces actions reposent sur 2 grand types de scripts/outils :

- Des scripts SQL permettant d'automatiser le calcul des descripteurs à l'échelle de l'USRA. Ces scripts s'exécutent sur PostgreSQL/PostGIS ;
- Un script R permettant de calculer les risques d'altération à l'échelle de l'USRA et de synthétiser les résultats aux échelles supérieures (TGH, ME DCE).

Ce document permet d'utiliser ou de comprendre l'organisation du code du dispositif PRHYMO. Par code, on entend ici à la fois les scripts R ou SQL permettant les calculs géomatiques et statistiques, mais aussi les scripts d'installation du serveur ainsi que les scripts de collecte des données d'entrée. On ne reprendra pas dans ce document le détail des scripts ou des procédures d'installation, tout étant tracé dans le code du dispositif.

L'arborescence du dispositif est la suivante :

- 0-install
- 1-collecte
- 2-traitements-sql
- 3-traitements-r
- 4-resultats

1- Spécifications à l'hébergement et à l'installation

Le dossier **0-installation** contient tous les éléments relatifs à l'installation du serveur.

1.1. Spécifications Serveur

L'ensemble des calculs se fait sur un serveur de calcul ayant les capacités minimales suivantes :

- 32Go RAM ;
- Pas de contraintes particulières sur le nombre de CPU ;
- 200Go SSD + 500Go SATA => la banque de données PostgreSQL utilise le SSD pour de meilleures performances et le SATA est utilisé pour le dépôt et la collecte de fichiers.

1.2. Spécifications Logicielles

Les scripts et procédures d'installation sont décrites dans le fichier README.md, notamment pour l'installation des logiciels et/ou packages suivants (entre parenthèses, version) :

- PostgreSQL 14 et PostGIS 3.2.1 + PgRouting ;
- Git, gitlab-runner pour le lancement des scripts depuis le dépôt de code ;
- Gdal (3.4.2) pour l'import de données spatialisées/géographiques ;
- R (4.2.0) avec les packages : sqldf(0.4.11), foreign(0.8.82), bnlearn(3.5), stringr(1.5.1), logging(0.10.108), igraph(2.0.1.1), gRbase(2.0.1), gRain(1.4.1).

2- Spécifications liées à l'architecture de la banque de données

La banque de données PRHYMO est structurée en plusieurs schémas respectant la nomenclature suivante :

- ref_* : pour les données de référence spécifiques aux réseaux hydrographiques (TGH, USRA, bassins versants) et les données qui en découlent (ex. zones tampons, remous liquides, USI) ;
- col_* : pour les données collectées (référentiel des obstacles à l'écoulement, occupation du sol, etc.) ;
- trt_* : pour les données intermédiaires issues des traitements ;
- bd_topo_* : pour la BD Topo ;
- stats_shared : pour la configuration du/des modèles bayésiens ;
- output_* : pour les sorties/résultats finaux

« * » peut prendre 6 valeurs représentant les grands territoires : Hexagone et 5 Départements et régions d'Outre-mer (DROM). Il y a donc la plupart du temps 1 schéma par territoire, sauf pour la configuration des modèles bayésiens car cette configuration est commune à tous les territoires.

Ce document ne décrit pas l'ensemble des tables (ex. tables intermédiaires de traitements, BD Topo ou données collectées), lesquelles sont tracées dans le code :

- Le premier paragraphe ci-dessous présente la structure des tables qui constituent les réseaux hydrographiques PRHYMO. Cette structure est commune à tous les territoires de manière à appliquer par la suite des traitements identiques sur cette base ;
- Le second paragraphe présente les tables générées à partir des USRA ;
- Le troisième paragraphe décrit les liens entre les réseaux hydrographiques, les masses d'eau DCE et les bassins versants ;
- Enfin, on trouvera en Annexe la liste des tables de la banque de données.

2.1. Structure des réseaux hydrographiques

Cette section décrit la structure initiale des TGH et USRA. La structure finale des couches que l'on peut diffuser aux utilisateurs résulte de la production de vues décrites au chapitre 3.4 - Résultats.

En entrée du processus de calcul des descripteurs et des paramètres élémentaires, la structure commune, ainsi que les attributs spécifiques à l'Hexagone ou aux DROM, sont en Figure 1. La plupart des traitements utilisent les attributs communs mais certains calculs de descripteurs peuvent exploiter une information renseignée par des attributs spécifiques. C'est par exemple le cas de la situation d'étiage naturel qui est calculée à partir de l'attribut q_etiage_n pour l'Hexagone et différemment pour les DROM.

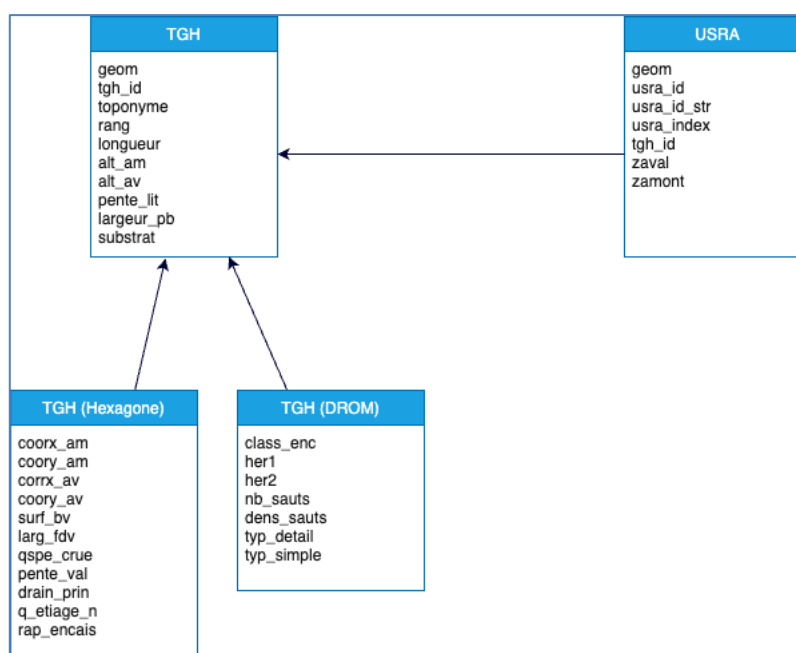


Figure 1 : structure initiale et liens entre tables USRA et TGH

TABLE REF_*.USRA :

Attribut	Définition
usra_id	Identifiant unique de l'USRA - format numérique
usra_id_str	Identifiant unique de l'USRA - format texte
tgh_id	Identifiant unique du Tronçon géomorphologiquement homogène (TGH)
usra_index	Numéro de l'USRA composant le TGH (numérotation de l'amont vers l'aval)
zamont	Altitude du nœud amont de l'USRA (en m)
zaval	Altitude du nœud aval de l'USRA (en m)

TABLE REF_*.TGH (Attributs communs utilisés dans les calculs) :

Attribut	Définition
tgh_id	Identifiant unique du Tronçon géomorphologiquement homogène (TGH) et clé primaire de la table des TGH
toponyme	Nom du cours d'eau (selon la BD Carthage)
rang	Rang de Strahler
longueur	Longueur du Tronçon géomorphologiquement homogène (en m)
alt_am	Altitude du nœud amont (en m NGF)
alt_av	Altitude du nœud aval (en m NGF)
pente_lit	Pente du lit (en %)
largeur_pb	Largeur à pleins bords théorique (en m)
substrat	Nature dominante du substrat en fond de vallée

2.2. Tables issues des USRA

Un certain nombre de tables sont calculées directement à partir de la table **USRA** (Figure 2). C'est le cas des Unités spatiales d'intégration (**USI**) pour le calcul de la rectitude et des tables **BUFFER**, lesquelles vont permettre d'effectuer les calculs dans des zones tampons déterminées autour des USRA. Comme expliqué au chapitre 3.2.2 - Calculs intermédiaires, chacune de ces tables est générée par une fonction SQL.

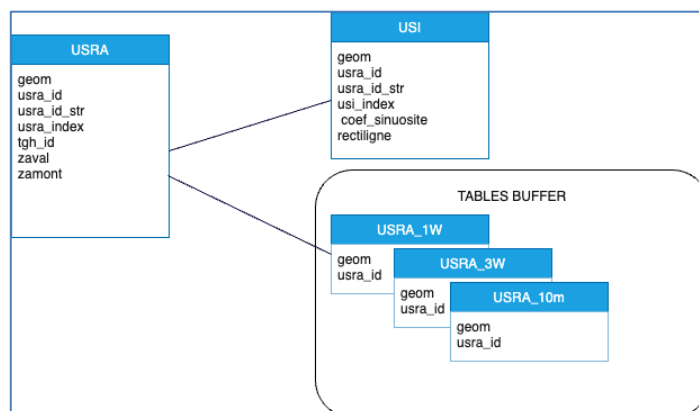


Figure 2 : structure initiale et liens entre tables USRA, USI et BUFFER

2.3. Liens avec les bassins versants et les masses d'eau

Le schéma ci-dessous (Figure 3) présente les liens entre les tables **TGH**, **USRA** et les tables contenant les masses d'eau (**ME**) ainsi que les bassins versants (**BV**).

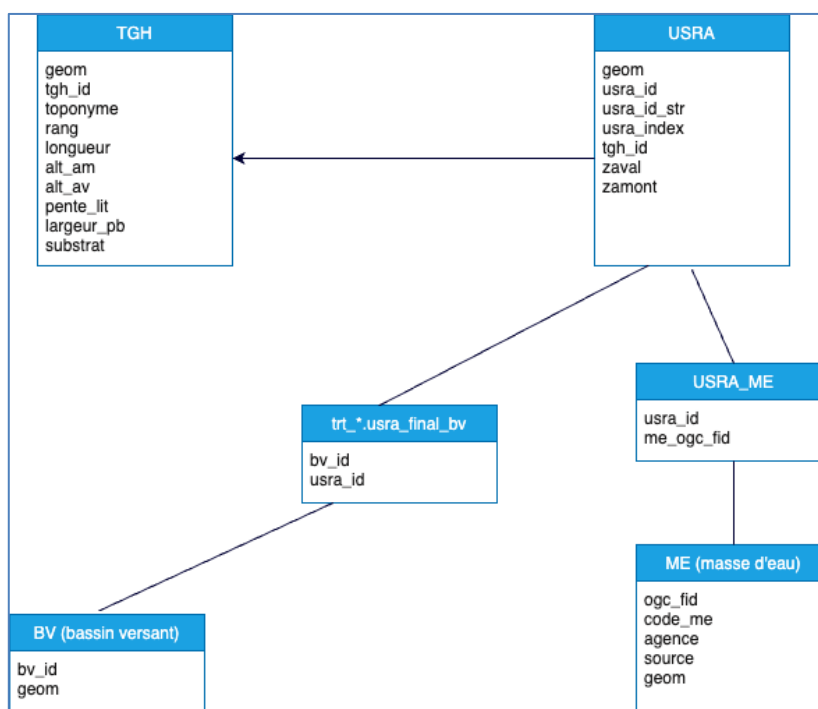


Figure 3 : liens entre les tables TGH, USRA, ME et BV

TABLE REF_*.ME :

Attribut	Définition
me_ogc_fid	Clé primaire de la table des masses d'eau (ME)
code_me	Code européen de la masse d'eau
agence	Code agence de la masse d'eau
source	Source de l'information

3- Spécifications pour l'exécution des calculs et traitements

3.1. Collecte

Le dossier **1-collecte** contient tous les scripts utilisés pour rapatrier les données d'entrée (Figure 4).

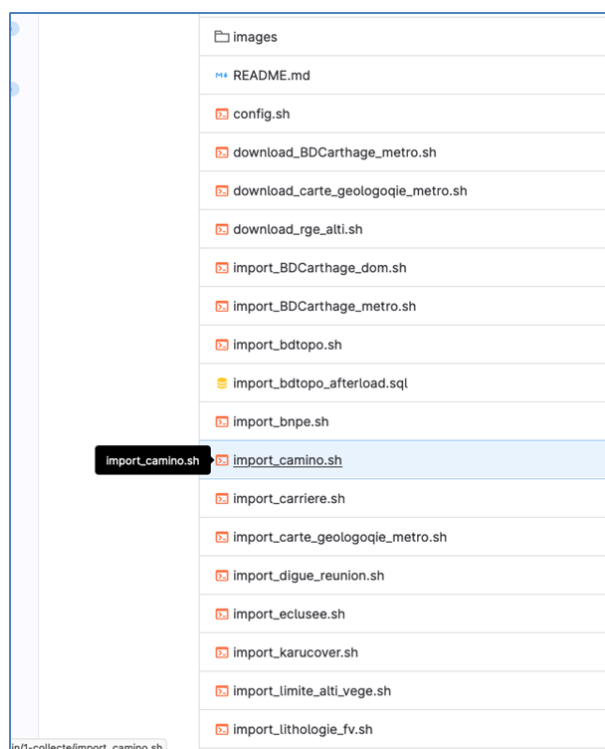


Figure 4 : scripts de collecte

import_bd_topo.sh par exemple télécharge la BD Topo et l'importe dans les bons schémas de la banque de données. **config.sh** permet de configurer l'accès à la banque de données et est utilisé par la plupart des scripts.

3.2. Traitements SQL

Le dossier **2-traitements-sql** contient les scripts :

- De calcul du réseau hydrographique en Hexagone : dans **ref_hydro/hexagone** ;
- De calcul de données intermédiaires nécessaires aux calculs d'indicateurs : liens amont/aval entre USRA, altitudes estimées des extrémités des USRA, liens avec les bassins versants, génération des unités spatiales d'intégration (USI) ;
- De calcul des descripteurs : dans **calc_descripteurs**.

3.2.1. Génération du réseau hydrographique en Hexagone

Le réseau hydrographique en Hexagone est réalisé en recalant les USRA de SYRAH-CE sur la géométrie de la BD Topo.

Le script **init_routing.sql** permet d'initialiser les tables de routing à partir des tronçons hydrographiques. Il utilise la fonction **pgr_createTopology** pour créer un réseau topologique dans lequel on pourra ensuite utiliser des fonctions de routing comme le calcul du plus court chemin entre 2 nœuds du réseau par exemple.

Il n'est à rejouer qu'en cas de modification de la BD Topo.

Le script principal (**rebuild_tgh.sql**) sert ensuite à créer les TGH et USRA PRHYMO à partir de la table **usra_work** qui est une copie des USRA SYRAH-CE pour lesquelles des modifications manuelles ont été apportées. Cette table peut être modifiée à l'avenir pour corriger des problèmes et améliorer à nouveau le réseau. Un cycle itératif de corrections a été réalisé pendant la construction du réseau PRHYMO et d'autres cycles pourraient être réalisés, sous réserve de relancer ensuite tous les scripts et étapes dépendantes.

Il s'agit d'un script complexe dont les principes sont les suivants :

- Projeter les points amont/aval de chaque USRA sur la BD Topo ;
- Utiliser le routing dans la BD Topo pour recréer les USRA à partir du réseau BD Topo ;
- Combler les discontinuités liées à ce changement de réseau ou présentes dans le réseau SYRAH-CE si possible ;
- Recréer des TGH en fusionnant les USRA ayant le même identifiant de tronçon.

Précisons que le comblement des discontinuités est impossible dans les cas suivants : (1) si la discontinuité est présente dans le réseau BD Topo ; (2) si les structures des réseaux SYRAH-CE et BD Topo sont incompatibles (ex. : confluence inexistante dans le nouveau réseau).

A la suite de ce script, un script shell (**2-traitements-sql/ref_hydro/hexagone/compute_tgh_alti.sh**) permet de calculer les altitudes nœuds amont/aval de chaque TGH en exploitant le MNT du RGE Alti.

Un script (**2-traitements-sql/ref_hydro/hexagone/publish_tgh_usra.sql**) permet ensuite de finaliser la génération de la couche TGH pour la mettre dans le bon schéma de référence avec les bons libellés de colonnes, index et contraintes.

Le schéma ci-dessous (Figure 5) représente une version simplifiée des tables et des scripts mis en jeu dans cette étape.

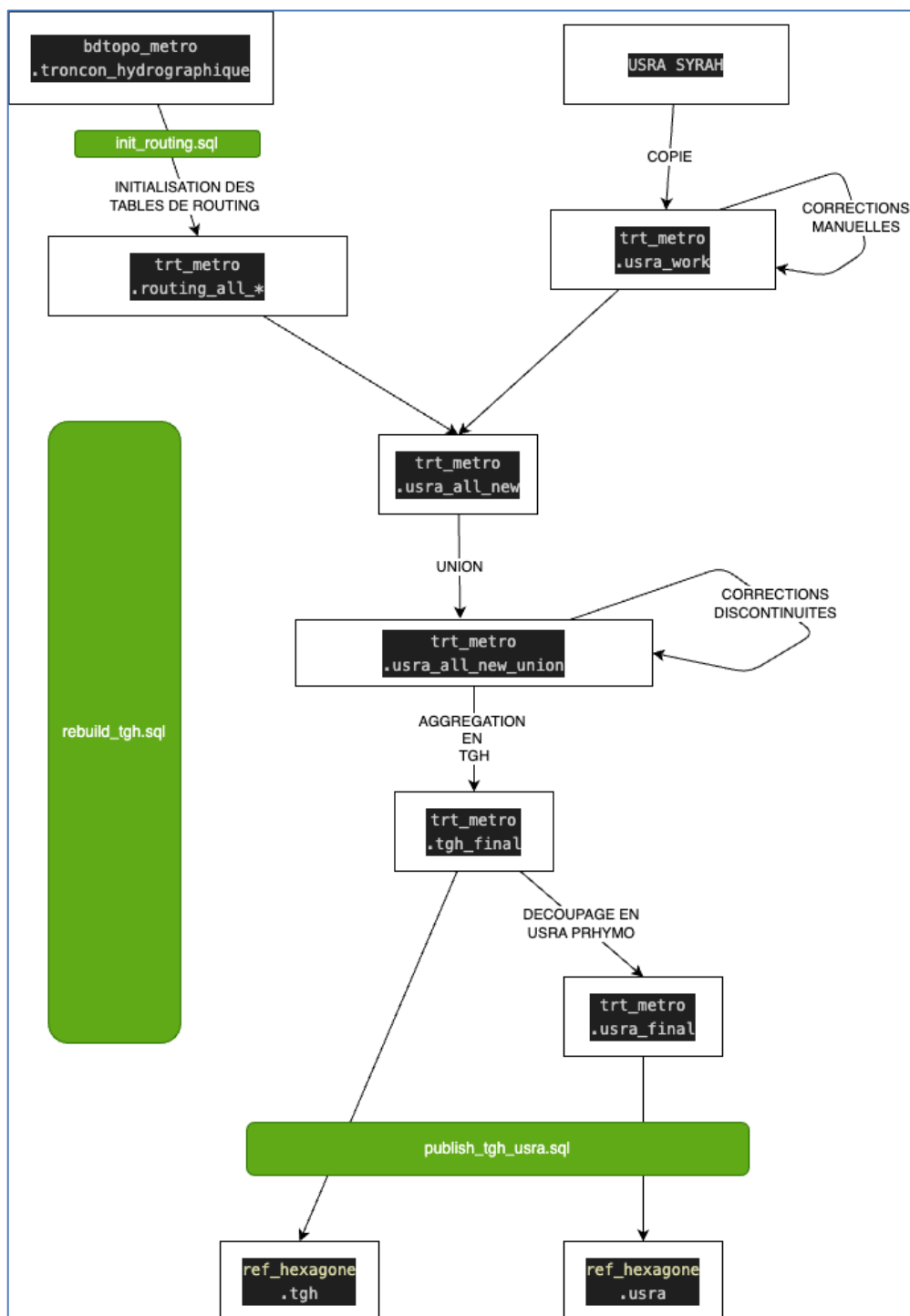


Figure 5 : version simplifiée des tables et des scripts mis en jeu dans cette étape

3.2.2. Calculs intermédiaires

Un certain nombre de tables intermédiaires sont créées pour permettre les calculs ultérieurs. Ces données sont générées par des fonctions SQL qui prennent en paramètre la plupart du temps :

- Le nom du schéma de traitement ;
- Le nom du schéma contenant les données de référence ;
- Le nom du schéma contenant les données collectées et parfois le nom du schéma contenant la BD Topo

Par exemple, pour calculer les USI en Hexagone, on va lancer la fonction suivante :

```
select calc_usi('trt_metro','ref_hexagone','col_hexagone');
```

Cette fonction est donc lancée sur les 6 territoires français avec les arguments associés et produit une table **USI** dans chaque schéma, contenant les données de référence.

```
select calc_usi('trt_martinique','ref_martinique','col_martinique');
select calc_usi('trt_guadeloupe','ref_guadeloupe','col_guadeloupe');
select calc_usi('trt_mayotte','ref_mayotte','col_mayotte');
select calc_usi('trt_reunion','ref_reunion','col_reunion');
select calc_usi('trt_guyane','ref_guyane','col_guyane');
select calc_usi('trt_metro','ref_hexagone','col_hexagone');
```

La source de cette fonction est dans le fichier **2-traitements-sql/ref_hydro/build_usi.sql** du dispositif PRHYMO mais peut aussi se retrouver dans la banque de données en parcourant les fonctions du schéma public avec **PG Admin** (Figure 6).

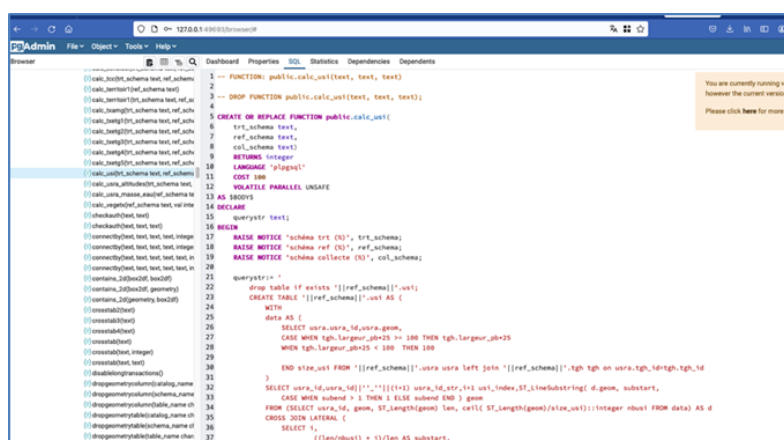


Figure 6 : visualisation du code d'une fonction dans PRHYMO via PG Admin

Voici les fonctions qui doivent être lancées pour ces calculs intermédiaires :

Fonctions ou enchainement de Fonctions	Étapes
calc_usi	Calcul des USI
calc_usra_masse_eau	Calcul de la table de jointure USRA<->Masse d'eau
calc_amont_aval_usra_final calc_amont_aval_tetes calc_amont_aval	Calcul du lien amont/aval entre les USRA
calc_bv_amont_aval	Calcul du lien amont/aval entre bassins versants
calc_usra_altitudes	Calcul de l'altitude estimée des nœuds des USRA a l'aide de la pente du tronçon
calc_roe_reseau_coulant calc_roe_reseau_coulant_points calc_roe_point_proj calc_roe_reseau_coulant_tetes calc_roe_reseau_coulant_amont_aval calc_remous_liquide	Calcul du remous liquide
<ul style="list-style-type: none"> • calc_buffer_usra_m : à 10, 30 et 60m • calc_buffer_usra_w : à 1, 3 et 12W • calc_buffer_diff: 3w-1w, 10m-1W, 30m-1W et 60m-1W 	Calcul des différents buffers autour des USRA

3.2.3. Calculs des descripteurs

De la même manière que pour les calculs intermédiaires, les descripteurs sont calculés par des fonctions SQL. A chaque descripteur correspond une fonction qu'il faut lancer avec les arguments correspondant au territoire ciblé. Chaque fonction va créer un nouvel attribut dans la table **ref_*.usra** pour le descripteur et éventuellement un second champ **<descripteur>_v** qui va stocker la valeur ayant permis de classer/discretiser le descripteur.

Par exemple, pour la rectitude on aura 2 attributs :

- rectitude_v = 25 par exemple
- rectitude = moyen par exemple

Exemple pour l'étiage en Guadeloupe :

```
select calc_etiage('trt_guadeloupe','ref_guadeloupe','ode971');
```

Pour chaque territoire, il y a un script permettant de lancer de manière globale le calcul de tous les descripteurs. Pour la Guyane, il s'agit par exemple du script suivant : **2-traitements-sql/calc-descripteurs/rebuild_all_guyane.sql**

3.3. Traitements R

3.3.1. Objectifs et principes

L'outil permet de calculer les risques d'altération de chaque paramètre élémentaire DCE à différentes échelles spatiales (USRA, TGH, ME) à partir, chacun, d'un modèle réseau de probabilités dites bayésiennes (*Bayesian Belief Network*) et des différentes caractéristiques (« descripteurs » évoqués précédemment) expliquant le paramètre étudié.

Dans un réseau bayésien, on dessine des arbres dont on distingue trois principaux éléments :

- Des nœuds **parents** qui représentent les descripteurs ;
- Des nœuds **enfants** qui sont le résultat de l'association d'un ou plusieurs parents ;
- Des **flèches** qui décrivent les relations causales entre les nœuds. Derrière chaque flèche, on décrit le lien de causalité par une distribution de probabilités conditionnelles. Ainsi, à chaque nœud enfant est associée une table de probabilités décrivant les probabilités du nœud enfant pour chaque combinaison de valeurs possibles pour les nœuds parents.

Le nœud enfant final représente le paramètre élémentaire DCE dans PRHYMO pour lequel on souhaite calculer les risques d'altération. Dans certains cas, il peut y avoir plusieurs nœuds intermédiaires, chacun étant à la fois parent et enfant. Par exemple, la pression sur les débits d'étiage est le nœud enfant de 2 descripteurs et un des 2 nœuds parents du paramètre élémentaire *Hydrologie – Quantité* (Figure 7).

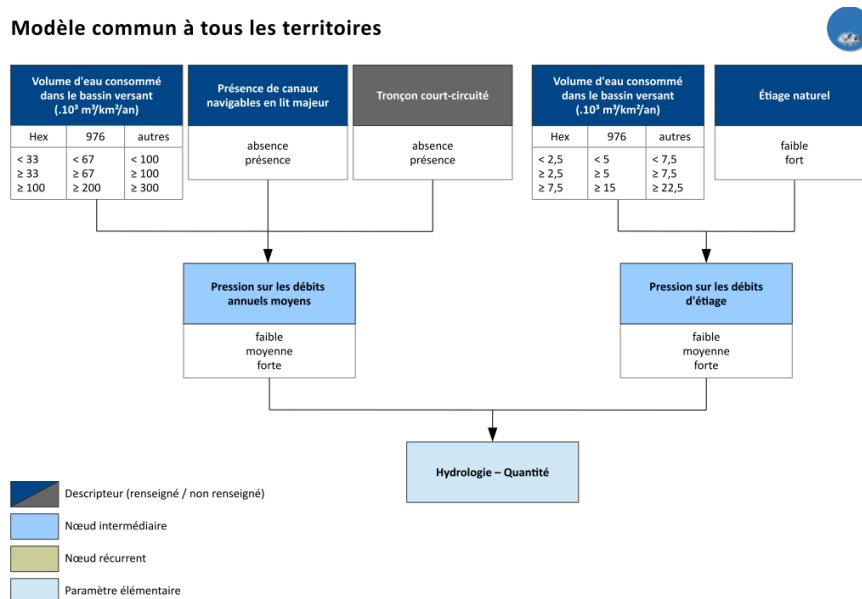


Figure 7 : exemple de l'arbre bayésien pour l'élément de qualité « Hydrologie - Quantité » (le document `prhymo_structure_arbres_bayesiens_062024_1_2.pdf` spécifie la structure de chaque arbre pour chaque élément de qualité et paramètre élémentaire DCE, Grosprêtre et Kreutzenberger 2024¹)

¹ Grosprêtre L., Kreutzenberger K., 2024. Plateforme Pressions et Risques d'impacts HYdroMORphologiques (PRHYMO) : structure des arbres bayésiens. Dynamique Hydro et Office Français de la Biodiversité, juin 2024.

3.3.2. Architecture du compartiment bayésien

Le compartiment bayésien de PRHYMO a été développé sous R (logiciel OpenSource et gratuit) et nécessite en particulier le package « *bnlearn* » pour les calculs bayésiens.

Le dossier « **3-traitements-r/prhymo** » contient les dossiers et scripts nécessaires à l'utilisation de ce compartiment. Il contient notamment un script (**prhymo.r**) et un fichier contenant des fonctions créées spécifiquement (**fonctions_bayésien_prhymo.r**). Le script est appelé lors du lancement des calculs. Il contient également un répertoire **shared**, décrit ci-dessous, qui se décline en autant de répertoires qu'il y a de modèles bayésiens (1 par paramètre élémentaire soit 10 au total ; Figure 8). Ces derniers définissent la configuration des modèles bayésiens dans le même formalisme que celui utilisé pour le Référentiel hydromorphologique ultramarin (RHUM).

Remarque importante concernant l'architecture du système :

Comme l'indiquent le rapport méthodologique et ses annexes, la structure conceptuelle des modèles bayésiens pour un paramètre élémentaire donné diffère selon les territoires. Néanmoins, **la structure informatique des modèles bayésiens est identique pour tous les territoires !** Cette structure unique facilite la modification des modèles et permet ainsi d'éviter certaines erreurs (une modification s'appliquera à tous les territoires). Par exemple, le modèle d'évaluation pour le paramètre *Continuité biologique pour les amphihalins* tient compte – structurellement parlant – de la présence de sites d'orpaillage légal sur tous les territoires. Mais la valeur par défaut du descripteur (ici absence d'orpaillage) ainsi que les filtres appliqués sur les données d'entrée (dont région = Guyane) confèrent logiquement une influence nulle à ce descripteur pour tous les territoires autre que la Guyane.

L'utilisateur peut changer les modèles, i.e. définir (1) les nœuds, (2) les relations entre chaque nœud et (3) les probabilités conditionnelles associées. Il faut pour cela créer plusieurs fichiers CSV qu'il faut placer dans une architecture rigoureuse de dossiers (ou modifier les fichiers CSV existants). On construit un dossier par paramètre élémentaire nommé « E_xxxxx ». Les modèles de chaque paramètre élémentaire sont regroupés dans un même dossier. Chaque dossier de paramètre élémentaire contient trois sous-dossiers : « **Arbre** », « **Descripteurs** », « **Enfants** » (Figure 9).

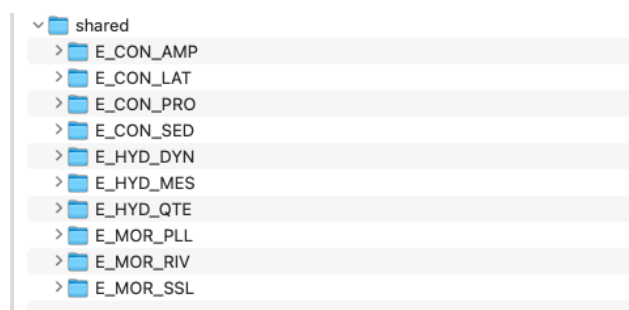


Figure 8 : structuration des dossiers bayésiens

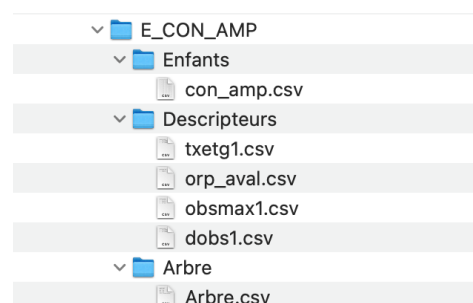


Figure 9 : structuration du dossier du paramètre élémentaire « con_amp » (exemple)

Dossier Arbre

Le dossier contient un seul fichier **Arbre.csv** (Figures 9 précédente et 10) : il décrit les nœuds et les relations.

Dans la colonne **PARAMETRE** sont listés les noms des variables. Pour les descripteurs, il est nécessaire de respecter les noms des attributs du shapefile et de les mettre en minuscule.

Une règle est appliquée au choix des noms des variables : le nombre de caractères est limité à 9, les espaces sont interdits et doivent être éventuellement remplacés par « _ ».

Dans la colonne **PARENTS**, mettre « **Descr** » est impératif s'il s'agit d'un descripteur. Par définition un descripteur n'a pas de parents. Pour les nœuds enfants, il est nécessaire d'indiquer les noms des nœuds parents séparés par deux points « : ».

Dans la dernière colonne **PELEM** enfin, il est nécessaire de mettre une petite croix « x » pour indiquer qu'il s'agit d'un paramètre élémentaire (nœud enfant final).

	A	B	C	
1	PARAMETRE	PARENTS	PELEM	
2	obsmax1	Descr		
3	txetg1	Descr		
4	dobs1	Descr		
5	orp_aval	Descr		
6	con_amp	orp_aval:obs	x	
7				
8				
9				
10				

Figure 10 : exemple d'une table descriptive d'un arbre (Arbre.csv)

Dossier Descripteurs

Le dossier contient autant de fichiers CSV que de descripteurs dans le modèle.

Chaque fichier CSV (Figure 11) prend le nom du descripteur en minuscule. Il définit chaque modalité (= valeur) des descripteurs ainsi que sa probabilité par défaut. Mais cette dernière n'est pas utilisée (si un descripteur n'est pas renseigné par les données collectées, la valeur définie par défaut lui est systématiquement attribuée avant cette étape du calcul).

	MODALITE	LABEL	PROB
1			
2	1	faible	0.33333333
3	2	moyen	0.33333333
4	3	fort	0.33333333

Figure 11 : exemple d'une table d'un descripteur (txetg1.csv)

Dans la colonne **MODALITE**, on indique un identifiant numérique entier. Cet identifiant sera utilisé dans les tables de probabilités conditionnelles (voir dossier **Enfants**).

Dans la colonne **LABEL**, on indique le nom de la modalité telle qu'elle est écrite dans le shapefile des USRA. Ainsi dans l'exemple, le shapefile doit contenir une colonne « txetg1 » contenant les modalités « faible », « moyen » et « fort ».

Dans la colonne **PROB**, on indique des probabilités (à dire d'expert) de rencontrer telle ou telle modalité du descripteur sur le territoire. Cette information est nécessaire au cas où le descripteur n'est pas alimenté par des données de connaissance des pressions dans la table d'attribut des USRA. Ces probabilités par défaut pourraient alors être utilisées mais ce n'est pas le cas car on attribue systématiquement une valeur aux descripteurs dans le cadre de PRHYMO (voir ci-dessus). Par exemple, pour le descripteur *TCC* (« présence d'un tronçon court-circuité »), comme nous n'avons pas de connaissances de cette pression sur le territoire, chaque USRA prendra la valeur par défaut soit « absence ». Par convention, nous répartissons les modalités de manière équitable (exemple 0.3333333 pour 3 modalités).

Les nombres dans la colonne **PROB** doivent être compris entre 0 et 1 et la somme doit faire strictement 1. Pour s'assurer que la somme fait 1, il est utile d'avoir recours à une formule type : « =1/3 » (cas dans Excel) pour afficher 0.3333333 au lieu de 0.33.

Dossier Enfants

Ce dossier contient les tables de probabilités conditionnelles des nœuds enfants.

Il y a autant de fichiers CSV que de nœuds enfants dans le modèle (Figure 9). Chaque fichier CSV prend le nom du nœud enfants en minuscule.

Les premières colonnes correspondent aux nœuds parents.

Les colonnes suivantes correspondent aux modalités du nœud enfant « X » (dans l'exemple suivant X1, X2, X3, X4, X5, correspondent respectivement à la probabilité d'une altération « Très faible », « Faible », « Moyenne », « Forte » ou « Très forte »).

La dernière colonne se nomme « **SOMME** » et doit être égale à 1. Elle permet de vérifier lors du remplissage des tables que la somme des probabilités est bien égale à 1 : elle est obligatoire.

Dans l'exemple suivant en ligne 6, si *contraint* = 1, *incision* = 2 et *txetg4* = 2 alors on aura 65% de risque que *con_lat* (paramètre élémentaire *Continuité latérale*) = 3 (moyen) et 35% de risque que *con_lat* = 4 (fort) (Figure 12).

con_lat									
	contraint	incision	txetg4	con_lat1	con_lat2	con_lat3	con_lat4	con_lat5	somme
1	1	1	1	1	0	0	0	0	1
2	1	1	2	0	0	1	0	0	1
3	1	2	1	0	0	1	0	0	1
4	2	1	1	0	0	1	0	0	1
5	1	2	2	0	0	0.65	0.35	0	1
6	2	1	2	0	0	0.65	0.35	0	1
7	2	2	1	0	0	0.65	0.35	0	1
8	2	2	2	0	0	0.35	0.65	0	1
9	1	1	3	0	0	0	1	0	1
10	1	3	1	0	0	0	1	0	1
11	3	1	1	0	0	0	1	0	1
12	1	2	3	0	0	0	0.75	0.25	1
13	2	1	3	0	0	0	0.75	0.25	1
14	1	3	2	0	0	0	0.75	0.25	1
15	2	3	1	0	0	0	0.75	0.25	1
16	3	1	2	0	0	0	0.75	0.25	1
17	3	2	1	0	0	0	0.75	0.25	1
18	2	2	3	0	0	0	0.5	0.5	1
19	2	3	2	0	0	0	0.5	0.5	1
20	3	2	2	0	0	0	0.5	0.5	1
21	1	3	3	0	0	0	0	1	1
22	3	1	3	0	0	0	0	1	1
23	3	3	1	0	0	0	0	1	1
24	2	3	3	0	0	0	0	1	1
25	3	2	3	0	0	0	0	1	1
26	3	3	2	0	0	0	0	1	1
27	3	3	3	0	0	0	0	1	1

Figure 12 : table des probabilités con_lat.csv

Remarques et préconisations importantes sur la mise en forme des dossiers et fichiers :

- Il faut scrupuleusement respecter les noms des descripteurs et enfants. Mettre des noms courts (moins de 9 caractères), sans espace, en minuscule ;
- Les fichiers doivent être au format CSV avec séparateurs point-virgule « ; ». Possibilité d'enregistrer un fichier Excel directement en CSV ;
- Il est nécessaire de vérifier que les sommes des probabilités pour chacune des combinaisons (en ligne) sont égales à 1 ;
- Pour enregistrer correctement un fichier CSV sous Excel, il faut cliquer sur la cellule A1, puis faire CTRL+A (sélectionner uniquement les cellules remplies, puis CTRL+S (enregistrer) ;
- Toutes les modalités doivent être décrites ;
- Si une cellule est vide, le calcul risque de renvoyer une erreur. Il faut donc saisir « 0 » si la probabilité est nulle.

3.3.3. Chargement des CSV en banque de données

Les CSV doivent être chargés dans la banque de données (dans le schéma **stats_shared**) pour être ensuite exploités par le script R. Le script **3-traitements-r/prhymo/load_descripteurs.sh** permet de réaliser cet import en banque de données.

3.3.4. Lancement des calculs bayésiens

Les calculs bayésiens sont réalisés par le script R **prhymo.r** qui doit être lancé par territoire :

```
Rscript prhymo.r <territoire> <debug : TRUE/FALSE>
```

```
Rscript prhymo.r martinique TRUE
```

Voici les commandes pour lancer le calcul sur l'ensemble des territoires :

```
cd 3-traitements-r/prhymo
mkdir OUTPUT
mkdir OUTPUT/mayotte
mkdir OUTPUT/reunion
mkdir OUTPUT/martinique
mkdir OUTPUT/guadeloupe
mkdir OUTPUT/hexagone
mkdir OUTPUT/guyane
Rscript prhymo.r guadeloupe TRUE
Rscript prhymo.r martinique TRUE
Rscript prhymo.r guyane TRUE
Rscript prhymo.r reunion TRUE
Rscript prhymo.r mayotte TRUE
Rscript prhymo.r hexagone TRUE
```

3.4. Résultats

L'exécution du fichier **4-resultats/resultats.sql** permet :

- De récupérer des valeurs d'attributs provenant des anciens TGH (informations produites par RHUM et SYRAH-CE et que l'on souhaite conserver) ;
- De mettre en forme les résultats via la production de vues prêtes à être exportées et diffusées aux utilisateurs finaux.

Les dernières tables ainsi produites (usra_synthese, tgh_synthese, me_synthese) ne reprennent pas tous les champs utiles aux calculs mais uniquement les champs utiles pour les utilisateurs finaux. Et certains champs changent de nom pour être plus compréhensibles par ces utilisateurs. La structure de ces tables est décrite dans le document **prhymo_dictionnaire_attributs_062024_1_2.pdf** (Grosprêtre et al. 2024)².

² Grosprêtre L., Despres D., Kreutzenberger K., 2024. Plateforme Pressions et Risques d'impacts HYdroMORphologiques (PRHYMO) : dictionnaire des attributs. Dynamique Hydro, Neogeo Technologies et Office Français de la Biodiversité, juin 2024.

3.5. Automatisation du lancement des calculs

Le fichier **gitlab-ci.yml** à la racine du dispositif PRHYMO permet de définir des tâches qui peuvent être lancées manuellement ou automatiquement depuis l'outil de gestion de code source en ligne (GitLab). On a donc un ensemble de tâches qui peuvent être relancées sur le serveur juste par un clic sur cette interface (Figure 13) et on peut accéder aux logs des scripts lancés de cette manière (Figure 14) :

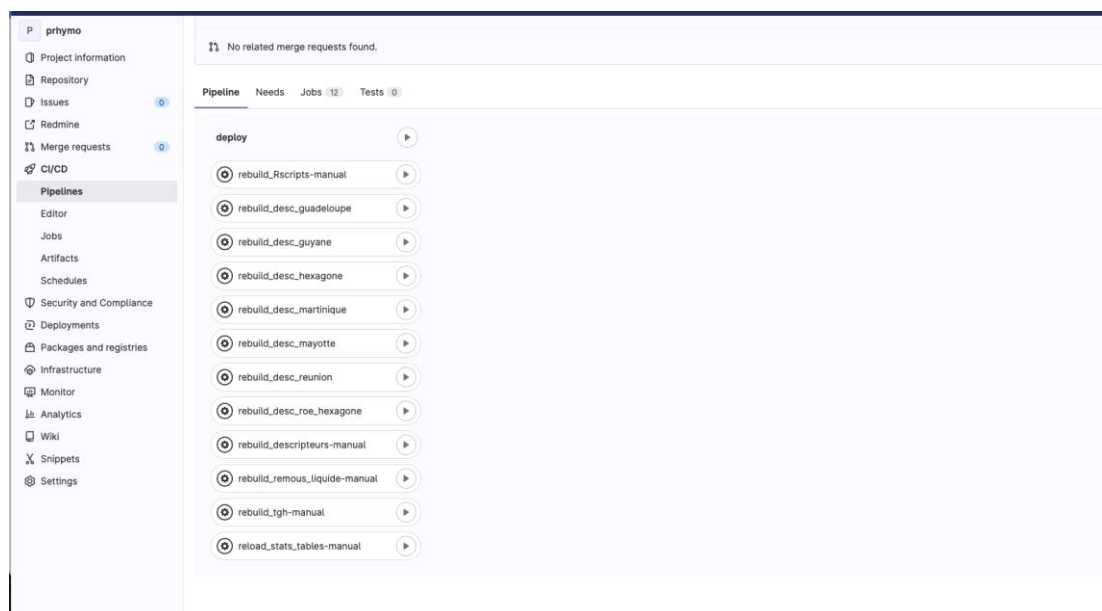


Figure 13: interface GitLab de lancement des scripts

L'intérêt de cette méthode est surtout d'éviter d'avoir à lancer manuellement des tâches sur le serveur et donc de garder une **traçabilité** complète des scripts lancés, de leurs paramètres *etc.*

En cas de mise à jour d'une probabilité du ou des modèles bayésiens dans un fichier CSV par exemple, il faudra lancer la tâche **reload_stats_tables-manual** pour recharger les CSV en base.

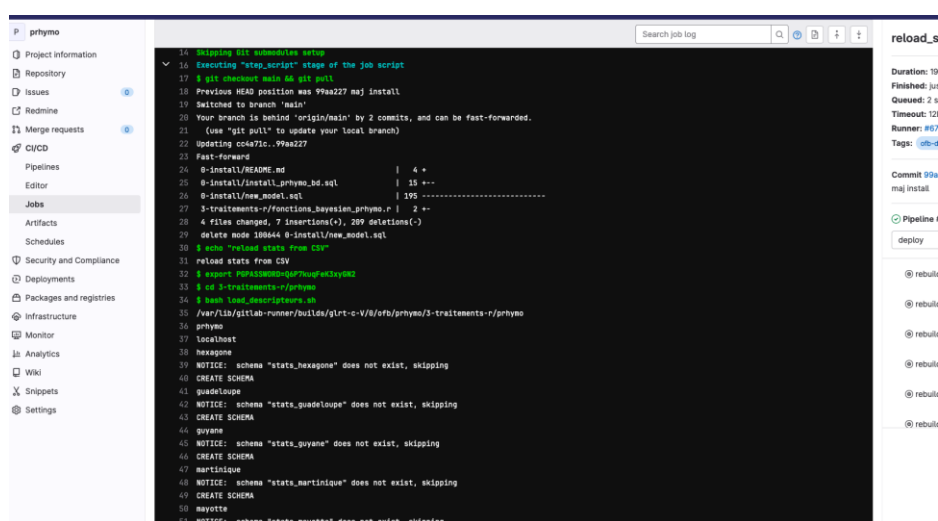


Figure 14 : affichage de la sortie d'un script sous GitLab

Pour relancer le recalage des USRA SYRAH-CE sur la BD Topo, il faut relancer **rebuild_tgh-manual**, qui enchaîne un certain nombre de commandes comme décrit plus haut :

```
.rebuild_tgh:
tags:
  - ofb-datastore
script:
  - git checkout main && git pull
  - echo "rebuild TGH from USRA_WORK"
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -c 'SELECT count(usra_simple_id) FROM trt_metro.usra_work'
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -f 2-traitements-sql/ref_hydro/hexagone/rebuild_tgh.sql
  - echo "calcul des altitudes"
  - sh 2-traitements-sql/ref_hydro/hexagone/compute_tgh_alti.sh
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -f 2-traitements-sql/ref_hydro/hexagone/publish_tgh_usra.sql
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -c "select calc_usi('trt_metro','ref_hexagone','col_hexagone');"
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -c "select
calc_amont_aval_usra_final('trt_metro','ref_hexagone','');"
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -c "select calc_amont_aval_tetes('trt_metro','ref_hexagone','');"
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -c "select calc_amont_aval('trt_metro','ref_hexagone','');"
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -c "select calc_bv_amont_aval('trt_metro','ref_hexagone','');"
  - psql -U prhymo_admin -h 127.0.0.1 -d prhymo -c "select calc_usra_altitudes('trt_metro','ref_hexagone','');"

```

Il y a des tâches pour relancer le calcul de tous les descripteurs pour chaque territoire et une tâche permettant de relancer les traitements R, et de finaliser ainsi les résultats.

4- Annexe

Le Tableau 1 présente la liste des tables de la banque de données PRHYMO

Type :

- SB : source brute dont les métadonnées correspondent aux sources originelles
- SR : source retravaillée dont les métadonnées sont propres et déduites des scripts
- TR : données temporaires générées par un traitement dans le but de réaliser un autre calcul
- OP : output dont les métadonnées sont strictement décrites par les documents qui accompagnent la diffusion des données de PRHYMO

Pour les tables de type SB ou SR, il est possible de se référer au document **prhyimo_sources_de_donnees_062024_1_2.xlsx (Grosprêtre, juin 2024)** pour obtenir les métadonnées complètes (description, provenance, etc.) des données d'entrée et comprendre les traitements qu'elles ont subis.

Tableau 1: liste des tables PRHYMO

Schéma	Table	Type	Contenu succinct
col_*	roe_enrichi	SB	Obstacles à l'écoulement (connaissances filtrées, retravaillées et enrichies). Table présente dans les 6 schémas territoriaux
col_*	retenues	SB	Retenues. Table présente dans les 6 schémas territoriaux
col_guyane	camino	SB	Cadastre minier numérique ouvert
col_guyane	piste	SB	Pistes agricoles et forestières en Guyane
col_guyane	carriere	SB	Exploitations minières en Guyane
col_guyane	occsol_2015	SB	Occupation du sol en 2015 sur la bande littorale de la Guyane
col_guadeloupe	karucover_2017	SB	Occupation du sol à grande échelle en 2 dimensions en Guadeloupe en 2017
col_martinique	occsol_2017	SB	Occupation du Sol à Grande Échelle de la Martinique en 2017
col_martinique	fond_beton	SB	RHUM - Tronçons à fonds bétonnés
col_martinique	curage	SB	RHUM - Tronçons curés
col_mayotte	occsol_2016	SB	Occupation du sol à Grande Échelle de Mayotte en 2016
col_reunion	occsol	SB	Occupation du sol 2021 à La Réunion
col_reunion	digue	SB	Digues de La Réunion
col_reunion	eclusee	SB	Ouvrages fonctionnant par éclusées à La Réunion
col_reunion	fond_beton	SB	RHUM - Tronçons à fonds bétonnés
col_hexagone	bnpe	SB	Prélèvements en eau période 2018 - 2020
col_hexagone	eclusee	SB	Ouvrages fonctionnant par éclusées en Hexagone

Schéma	Table	Type	Contenu succinct
col_hexagone	limite_alti_vege	SB	Limites altitudinales de la végétation arborée dans les massifs montagneux en Hexagone
col_hexagone	lithologie_fv	SB	Fonds de vallées alluviaux en Hexagone
col_hexagone	lithologie_fv_sub	SR	Subdivision de la couche lithologie_fv
col_hexagone	oso2021	SB	Occupation du sol en Hexagone en 2021
col_hexagone	sensibilite_erosion_sols	SB	Sensibilité des sols à l'érosion en Hexagone
bdtopo_*	*	SB	BD Topo
bdcarthage_*	*	SB	BD Carthage
ref_*	bv	SB	Bassins versants. Table présente dans les 6 schémas territoriaux
ref_*	me	SB	Masses d'eau. Table présente dans les 6 schémas territoriaux
ref_*	remous_liquide	OP	Remous liquide par obstacle du ROE enrichi. Table présente dans les 6 schémas territoriaux
ref_*	remous_liquide_union	OP	Union des géométries des remous liquides. Table présente dans les 6 schémas territoriaux
ref_*	tgh	SB	Tronçons géomorphologiquement homogènes. Table présente dans les 6 schémas territoriaux
ref_*	usi	OP	Découpage des USRA en USI pour pouvoir calculer la rectitude de l'USRA. Table présente dans les 6 schémas territoriaux
ref_*	usra	SB + OP	La géométrie des USRA est une source brute (fournie par Dynamique Hydro) pour Mayotte, Réunion, Guyane et Guadeloupe. Table présente dans les 6 schémas territoriaux
ref_*	usra_1w	OP	Buffer 1W appliqué aux USRA. Table présente dans les 6 schémas territoriaux
ref_*	usra_3w	OP	Buffer 3W appliqué aux USRA. Table présente dans les 6 schémas territoriaux
ref_*	usra_3w_moins_1w	OP	Buffer 3W appliqué aux USRA auquel est retiré le buffer 1W. Table présente dans les 6 schémas territoriaux
ref_*	usra_10m	OP	Buffer de 10m appliqué aux USRA. Table présente dans les 6 schémas territoriaux
ref_*	usra_10m_moins_1w	OP	Buffer de 10m appliqué aux USRA auquel est retiré le buffer 1W. Table présente dans les 6 schémas territoriaux
ref_*	usra_12w	OP	Buffer 12W appliqué aux USRA. Table présente dans les 6 schémas territoriaux
ref_*	usra_30m	OP	Buffer de 30m appliqué aux USRA. Table présente dans les 6 schémas territoriaux

Schéma	Table	Type	Contenu succinct
ref_*	usra_30m_moins_1w	OP	Buffer de 30m appliqué aux USRA auquel est retiré le buffer 1W. Table présente dans les 6 schémas territoriaux
ref_*	usra_60m	OP	Buffer de 60m appliqué aux USRA. Table présente dans les 6 schémas territoriaux
ref_*	usra_60m_moins_1w	OP	Buffer de 60m appliqué aux USRA auquel est retiré le buffer 1W. Table présente dans les 6 schémas territoriaux
ref_*	usra_me	OP	Table de jointure entre USRA et masses d'eau. Table présente dans les 6 schémas territoriaux
stats_shared	*		Modèles bayésiens : ce sont les CSV décrits au chapitre 0 - Architecture du compartiment bayésien qui sont chargés dans les tables de ce schéma
output_*	me_synth_rst	OP	Sortie du script R (prhymo.r), résultats bruts. Table présente dans les 6 schémas territoriaux
output_*	tgh_synth_rst	OP	Sortie du script R (prhymo.r), résultats bruts. Table présente dans les 6 schémas territoriaux
output_*	usra_synth_rst	OP	Sortie du script R (prhymo.r), résultats bruts. Table présente dans les 6 schémas territoriaux
output_*	me_synthese	OP	Résultats finaux – agrégation échelle masses d'eau. Table présente dans les 6 schémas territoriaux
output_*	tgh_synthese	OP	Résultats finaux – agrégation échelle TGH. Table présente dans les 6 schémas territoriaux
output_*	usra_synthese	OP	Résultats finaux échelle USRA. Table présente dans les 6 schémas territoriaux
trt_*	agri_intens	TR	Donnée temporaire générée par la fonction calc_agribv pour extraire de l'occupation du sol les zones agricoles
trt_*	amenagement	TR	Donnée temporaire générée par la fonction calc_txamg pour rassembler les données d'aménagement provenant de la BD Topo et de l'occupation du sol
trt_*	amenagement_sub	TR	Subdivision des géométries de la table amenagement pour accélérer les calculs
trt_*	gd_barrage	TR	Donnée temporaire générée par la fonction calc_intercepte pour extraire les grands barrages du ROE
trt_*	imper	TR	Donnée temporaire générée par la fonction calc_imper pour extraire de l'occupation du sol les zones imperméables
trt_*	reseau_coulant	TR	Découpage des USRA aux confluences. Donnée temporaire générée par la fonction

			calc_roe_reseau_coulant pour le calcul du remous liquide
Schéma	Table	Type	Contenu succinct
trt_*	reseau_coulant_amont_aval	TR	Table de lien amont/aval du réseau coulant. Donnée temporaire générée par la fonction calc_roe_reseau_coulant_amont_aval
trt_*	reseau_coulant_amonts	TR	Table de lien entre un tronçon du réseau coulant et ses tronçons amonts. Donnée temporaire générée par la fonction calc_roe_reseau_coulant_amont_aval
trt_*	reseau_coulant_tetes	TR	Têtes de réseau du réseau coulant (tronçons les plus amonts uniquement). Donnée temporaire générée par la fonction calc_roe_reseau_coulant_texte
trt_*	roe_reseau_coulant	TR	Projection des points du ROE sur le reseau coulant. Donnée temporaire générée par la fonction calc_roe_reseau_coulant_points
trt_*	usra_final	TR	Copie de la table ref_*.usra pour les calculs de lien amont/aval entre USRA. Donnée temporaire générée par la fonction calc_amont_aval_usra_final
trt_*	usra_final_amont	TR	Table de lien entre une USRA et ses USRA amonts. Donnée temporaire générée par la fonction calc_amont_aval_usra_final
trt_*	usra_final_amont_aval	TR	Table permettant de faire le lien amont/aval entre les USRA. Donnée temporaire générée par la fonction calc_amont_aval
trt_*	usra_final_amonts	TR	Dernières USRA en têtes de réseau amont (donc tronçons les plus amonts uniquement) Donnée temporaire générée par la fonction calc_amont_aval_tetes
trt_*	usra_final_bv	TR	Table de lien entre les USRA et les bassins versants. Donnée temporaire générée par la fonction calc_bv_amont_aval
trt_*	usra_final_bv_amont	TR	Table de liens amont/aval entre les bassins versants. Donnée temporaire générée par la fonction calc_bv_amont_aval
trt_*	usra_final_confluences	TR	Calcul des points de confluence pour le découpage des USRA en réseau coulant. Donnée temporaire générée par la fonction calc_roe_reseau_coulant
trt_*	usra_with_roe	TR	Calcul des USRA ayant un obstacle proche. Donnée temporaire générée par la fonction calc_obsmax2
trt_metro	usra_work	SR	USRA issues de SYRAH-CE modifiées manuellement via QGIS. Entrée des scripts de recalage sur la BD Topo
trt_metro	bv_haie	TR	Somme de longueur de haies par bassin versant. Donnée temporaire générée par la fonction calc_haies

Le Tableau 2 présente les tables importées des dispositifs RHUM et SYRAH-CE ayant servi à alimenter les données d'entrée du dispositif PRHYMO (i-e, elles ne sont pas produites dans le cadre de PRHYMO).

Tableau 2 : liste des tables importées des dispositifs RHUM et SYRAH-CE

Schéma	Table	Type	Contenu succinct
ode972	digue	SB	RHUM - Digue
ode971	deb_etiage	SB	RHUM - Zones d'étiage naturellement fort/faible
ode972	deb_etiage	SB	RHUM - Zones d'étiage naturellement fort/faible
ode974	deb_etiage	SB	RHUM - Zones d'étiage naturellement fort/faible
ode976	deb_etiage	SB	RHUM - Zones d'étiage naturellement fort/faible
rhum	*	SB	Résultats finaux à l'échelle TGH du dispositif RHUM utilisés pour la comparaison des résultats avec les résultats PRHYMO
syrah	usra	SB	SYRAH-CE : USRA – Unités spatiales de recueil et d'analyse
syrah	usra_2017_donnees_entree_bayisien_v2	SB	SYRAH-CE : descripteurs des USRA SYRAH-CE, l'information drain principal et q_etiage_n (situation d'étiage naturel) est récupérée ici et reportée dans les TGH PRHYMO
syrah	usra_res	SB	Résultats finaux à l'échelle USRA du dispositif SYRAH-CE utilisés pour la comparaison des résultats avec les résultats PRHYMO
syrah	tgh	SB	SYRAH-CE : TGH – Tronçons géomorphologiquement homogènes.
syrah	tgh_rst	SB	Résultats finaux à l'échelle TGH du dispositif SYRAH-CE utilisés pour la comparaison des résultats avec les résultats PRHYMO