

Gestion du flux de données en provenance du GTS ou SMT (Système Mondial des Télécommunications)

Mohamed Arab Benamara^{1*}, Idir DEHMOUS^{1*}, Mohamed BOUZGHAIA^{1*}

Abstract

La mesure ou l'observation est très importante ou même capitale dans le domaine de la météorologie, elle est utile pour les deux branches de cette dernière, la première branche est la climatologie qui consiste essentiellement à archiver des observations pour les statistiques faites en prévision climatique (modèles statistiques et climatologiques), la deuxième branche est la prévision numérique du temps (PNT) qui consiste à faire tourner des modèles numériques, ces derniers sont initialisés par la combinaison d'un état atmosphérique connu (prévision à courte échéance) et les observations valides au moment du lancement de la prévision.

Au niveau de l'ONM, la collecte des messages d'observations météorologiques est assurée par le cntm (centre national des télécommunications météorologiques), en qualité de centre national de de centre régional des télécommunications (CRT). Les services demandeurs de ces messages sont variés. Par ailleurs, afin d'assurer une exploitation optimale des messages, leurs transmissions via le Système Mondial des Télécommunications de l'OMM se fait d'une manière automatique au moment opportun.

Pour la période de l'étude du 01 au 31 mars 2018, le nombre total de messages textes METAR, SYNOP, TEMP, TAF ET AMDAR inventoriés a dépassé le million, soit 1009419. Pour les statistiques des retards, le retard moyen global des messages reçus était 3.38 heures depuis l'heure de l'émission jusqu'à l'heure de la réception. Pour l'inventaire des fichiers binaires, les retards moyens dans l'émission/réception varient de 1,33 h à 7,5 h.

Keywords

Message d'observation, CRT, SMT, classification des messages

¹ Office national de la météorologie, Dar El Beida, Alger

*Correspondant: m.benamara@meteo.dz

Contents

Introduction	1
1 Méthodologie	1
2 Classification des Messages selon leurs Formats et contenus	2
3 Statistiques des Messages	3
3.1 Inventaire des messages	3
3.2 Réception des messages	3
4 Classement des messages selon l'entête WMO	3
4.1 4.1 Vers Une Base de données Interrogeable . .	3
5 Conclusion	4
References	4

Les observations météorologiques sont variées et diversifiées à ce titre leur classification s'avère difficile et compliquée. Dans notre service qu'est la PNT (prévision numérique du temps), le besoin en données venant des réseaux national et mondial est primordial pour l'aboutissement de nos programmes, de l'initialisation (adaptation dynamique) à l'assimilation (fabrication de l'état initial) jusqu'au contrôle, la mesure ou l'observation météorologique est vitale. De ce fait, avoir une information sur le flux, les types de données et leurs disponibilités est très important.

Dans cette étude, plusieurs approches ont été réalisées pour identifier les types et les formats des messages ainsi que

leurs contenus. La période de l'étude est basée sur les données de la période du 01 au 31 mars de l'année 2018.

Dans notre démarche nous avons commencé par la première classification qui consistait à classer ou séparer le flux de données en fichiers textes et en fichiers binaires. Les fichiers en format texte ont été identifiés en repérant les mots clés de chaque type de message connu (AAXX pour les synops terrestres, METAR pour les metars, AMDAR pour les amdars, etc). L'autre catégorie des fichier qui arrive via le SMT est de format binaire. Ces derniers sont repérés grâce aux mots clés GRIB [1] et BUFR [2] qui apparaissent en les lisant comme des fichiers textes (en utilisant la commande « vi » ou « cat » sous unix). La deuxième classification ou l'autre démarche qu'on a réalisé dans le but de trouver la méthode optimale était, d'adopter ou de suivre les tables [3] mises au point par l'Organisation Météorologique Mondiale (OMM) qui consiste en un système de nomination symbolique des fichiers échangé via le réseau GTS [4], qui est conçu principalement pour fournir un moyen par lequel les messages échangés peuvent être reconnus par le système de gestion des données et garantir l'unicité de chaque bulletin lors des processus d'émission réception.

1. Méthodologie

L'exploitation des messages météorologiques est multiple. Les services demandeurs de ces mêmes messages sont aussi

variés. Par ailleurs, afin d'assurer une exploitation optimale des messages, leurs transmissions à qui de droit, devrait se faire d'une manière automatique au moment opportun. La problématique est la mise en place d'un modèle de classification des messages météorologiques qui assurera une exploitation optimale. Les étapes à suivre sont les suivantes:

- a- Identification des messages météorologiques
- b- Tri et classification des messages météorologiques
- c- Redirection automatique des messages météorologiques

2. Classification des Messages selon leurs Formats et contenus

Pour accomplir ce travail, nous avons procédé par deux approches. La première approche est la classification des observations selon leurs formats.

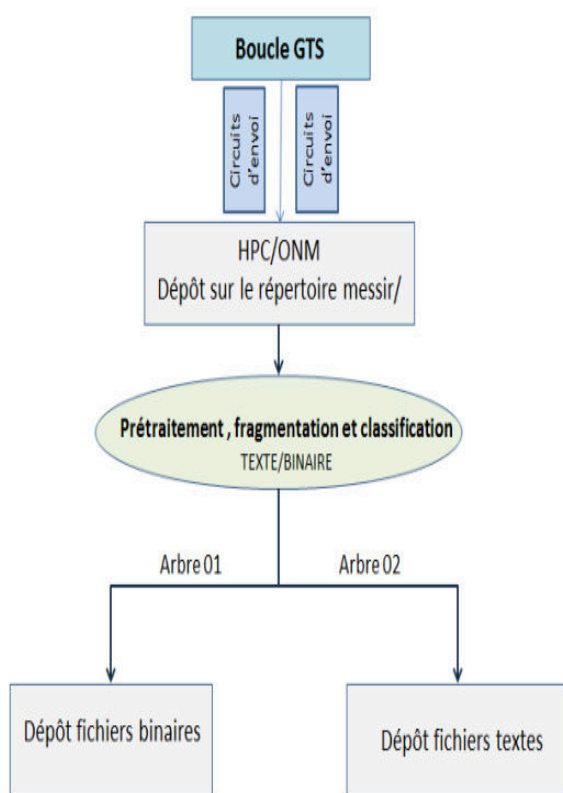


Figure 1. Organigramme de cheminement des données

Ce schéma nous montre le cheminement des données, de la boucle GTS jusqu'à leur séparations selon le type de format texte ou binaire à leurs aboutissements dans leurs dépôts respectifs. Cette classification consistait à classer ou séparer le flux de données en fichiers textes et en fichiers binaires. Ensuite on est passé à la reconnaissance des contenus des fichiers en format texte ou binaire en les repérant

par les mots clés de chaque type de message connu.

Exemple d'un message SYNOP :

```

SMAL20 DAMM 301200 RRA
AAXX 30124
60585 12958 00818 10199 20005 40223 58012 60001
333 0//00=
    
```

Exemple d'un message METAR

```

SAAL33 DAAA 301900
METAR DAAE 301900Z 19006KT CAVOK 09/07 Q1028=
    
```

Exemple d'un message AMDAR

```

UDAF01 FAPR 081638
AMDAR 0816
ASC AFZA49 3404S 01837E 081635 F022 PS135 196/013
ASC AFZA49 3404S 01837E 081635 F024 PS124 202/010
    
```

De même pour les binaires qui sont repérés grâce aux mots clés GRIB et BUFR qui apparaissent en les lisant comme des fichiers textes (en utilisant la commande « vi » ou « cat » sous unix) :

Exemple des messages de types GRIB

```

0002403700^A^M^M
106^M^M
HTXX99 LFPW 280600^M^M
GRIB^@]Â^A^@^@ (^AUÓÿ<80>^Kd
    
```

Exemple des messages de type BUFR

```

0000039500^A^M^M
966^M^M
ISMA02 EKMI 251200 RRC^M^M
BUFR^@^Ad^D^@^@^V^@^@^@^@^
    
```

Les fichiers binaires sont classés en fichiers GRIB et en fichiers BUFR en un premier temps. Puis on les classera selon leurs contenus. Les fichiers en format binaire principalement les GRIB, ont été classés en lisant le « header » ou entête. Le header nous renseigne sur le centre d'émission ou d'élaboration, de la date de sa validation, du paramètre dont il s'agit, du niveau (altitude ou couche), de la résolution horizontale et du domaine ou de la zone couverte par le champs, l'utilitaire utilisé est « wgrib » développé par le NCEP (National Center for Environment Prediction).

Les autres fichiers binaires traités sont en format BUFR, Utilisé principalement pour l'échange des données observations telles que, les données satellites. L'utilitaire utilisé pour leur décodage est « Bufr_dump » développé par l'ECMWF (European Centre for Medium Weather Forecast)

```

1-4  identifieur = BUFR
1-3  sectionLength = 22
4     masterTableNumber = 0
7-8  bufrHeaderSubCentre = 0
9     updateSequenceNumber = 0
10    sectionFlags = 0 [00000000]
11    dataCategory = 4
12    internationalDataSubCategory = 0
13    dataSubCategory = 255
    
```

3. Statistiques des Messages

Inventaire des messages

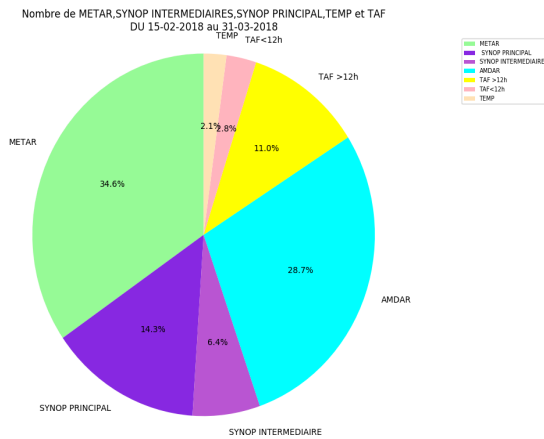


Figure 2. Inventaire des messages

La figure 2 nous montre la part de chaque type de fichier dans l'inventaire que l'on a réalisé durant la période de l'étude.

Réception des messages

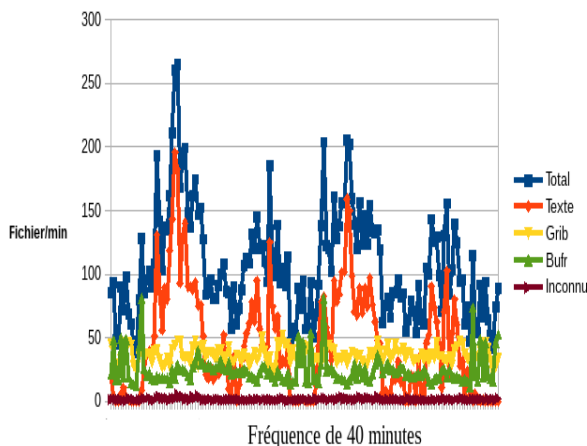


Figure 4. flux de réception des messages

La figure 4, montre le flux de réception important de tous les messages confondus par une fréquence de 40 minutes (texte, bufr, grib, autres). Ci-dessous nous afficherons quelques statistiques des fichiers grib et bufr.

Indicatif	Nombre	Centre	Résolution
HEXW98	76	DAMM	0,08
HEXX00	1522	LFPW	0,50
HEXZ98	38	ECMF	0,50
HEYB98	38	EGRR	1.00

Indicatif	Nombre	Centre	Catégorie
IOBA01	1972	MADRID	SURF-SEA
ISAI01	6352	BRASILIA	SURF-LAND
ISBC01	86	TOKYO	Radar
ISCA01	7	DUBLIN	SURF-LAND

4. Classement des messages selon l'entête WMO

Le système de nomenclature symbolique basé sur des tables mises au point par l'OMM et utilisé par le GTS permettant de :

- a- Codification commune à tous les types de données.
- b- Identification rapide grâce aux tables (tables WMO N°386, Vol I)
- c- Garantir l'unicité de chaque bulletin lors des processus d'émission réception.

La forme générale des noms des fichiers est donnée par une forme symbolique : « T1T2A1A2ii_CCCC_YYGGgg_(BBB) »

Symbole	Signification
T1T2	Forme générale du bulletin et Type de données
A1A2	Emplacement géographique du bulletin
ii	Indicatifs numériques
CCCC	Indicateur international du centre qui est à d'origine du message
YYGGgg	Indicateur Temporel
BBB	Indicateur d'addition, de correction ou d'amendement

A travers ce tableau nous avons essayé de montrer, les motifs qui font la distinction parmi les différents types de messages et aussi en les inventoriant.

Classement Synthétique Par Motif (Résultats)

Type	Sous-Type	Motif	Nombre /jour
GRIB	Arpege	^(H Y).*.LFPW.*\$	21693
	IFS	^(H Y).*.EGRR.*\$	13676
	EMet	^(H Y).*.ECMF.*\$	838
	Aladin	^(H Y).*.DAMM.*\$	1921
	LEMet	^(H Y).*.RWBC.*\$	5127
	SFS	^(H Y).*.NCEP.*\$	0
ImageSat	Oceano	^O.*\$	0
	Other	^(H Y).*\$	380
	All	^(E T).*\$	0
	Analyses	^A.*\$	2
	Chim at	^C.*\$	18
	Provisions	^P(F C).*\$	3423
	Other	^O.*\$	13
	GRID	^G.*\$	37
	Obs-Bufr	^S.*\$	0
	Obs-Bufr	Surface	^SS.*\$
Obs-Bufr	Altitude	^SU.*\$	11660
Obs-Bufr	Other	^S.*\$	4403
Pres-Bufr	All	^P.*\$	66
CREX	All	^R.*\$	21
NMIL	All	^L.*\$	0
Notices	All	^N.*\$	0
Pictorial	All	^(P Q).*\$	225
Observation de surface	ME TAR	^SA.*\$	8806
	RADOB	^S(B C D).*\$	0
	SYN OP	^S(F MIN).*\$	6023
	BUOY	^SS.*\$	0
	WAVE OB	^SW.*\$	0
Observations d'altitude	Other	^O.*\$	3523
	AIREP	^UA.*\$	448
	AMDK AR	^UD.*\$	6763
	TEMP	^(S E I E I E I F S).*\$	2061
	PILOT	^U(P Q H Q I Y).*\$	583
Alertes	Other	^U.*\$	0
	All	^W.*\$	97
Other	Other	^.*\$	0

Figure 5. inventaire et motifs synthétiques

4.1 Vers Une Base de données Interrogeable

Afin de rendre notre travail exploitable nous sommes passé à l'étape de création de base de données interrogeable dont la démarche est ci dessous : Aperçu de la base de données Base

de données SQLite (Flexibilité + légèreté) Identification des données par :

- Entête
- Date de création
- Date de référence
- Numéros d'ordre

Exemple de Requête

Pour interroger la base de données nous avons suivi les étapes suivantes :

- Commande : GtsGet (Accessible sur le HPC de l'ONM)
- Arguments d'entrée :
 - Date de Début : 3 mois avant la date de la requête
 - Date de Fin : instant de la requete
 - Motif de correspondance de l'entête : "^AFT.*\$" (correspondant à tous les TAF courts)
 - Dossier de dépôt des données : "data"

5. Conclusion

Dans le but, d'identifier tous les messages reçus via le SMT, nous avons entrepris dans un premier temps une démarche qui consistait à classer les messages selon leurs types d'écriture et selon leurs contenus repérés par les mots clés qui différencient les messages. Nous avons terminé par un inventaire des messages répertoriés et nous avons donné les statistiques sur le type de messages reçus et les retards entre l'émission et la réception. Dans une seconde étape nous avons opté pour un classement selon les tables établies par l'OMM. Ces tables permettent d'identifier tous les messages quelque soient leurs formats et contenus. Pour la période d'étude, le nombre total de tous les messages textes (METAR, SYNOP, TEMP, TAF et AMDAR) obtenu est de l'ordre de 1009419.

Les nombres de ces messages en pourcentage étaient de 26.86% pour les metar, 16.05% pour les synop, 22.27% pour les amdar, 1.66% pour les Temp, et 10.71% pour les Taf. En ce qui concerne les statistiques des retards, la moyenne des retards de tous les messages reçus était de 3.38 heures depuis l'heure de l'émission jusqu'à l'heure de la réception. L'inventaire fait dans le cadre de cette étude montre que le nombre des fichiers binaires est plus importants par rapport aux fichiers textes. Ces fichiers contiennent des champs aux points de grille des différents paramètres, pression, humidité relative, précipitation, température, vent, vitesse verticale, rayonnement, divergence, houle et vague en format grib et d'autres en format bufr contiennent les observations en surface, en altitude et des mesures des propriétés océanographiques et limnographiques dont les retards moyens dans l'émission/réception varie entre 1,33 h à 7,5 h.

References

- [1] Ncep wmo grib2 documentation.
- [2] Shawn R Smith, Eric Freeman, Sandra J Lubker, Scott D Woodruff, Steven J Worley, William E Angel, Dave I Berry, Philip Brohan, Zaihua Ji, and Elizabeth C Kent. The international maritime meteorological archive (imma) format. *disponible en la página web: <http://icoads.noaa.gov/e-doc/imma>*, 3, 2016.
- [3] BMO WMO, OMM. *CATALOGUE OF METEOROLOGICAL BULLETINS*. Number 9. 2012.
- [4] WMO. *Manual on the Global Telecommunication System*. Number 386. 8 edition, 2009.

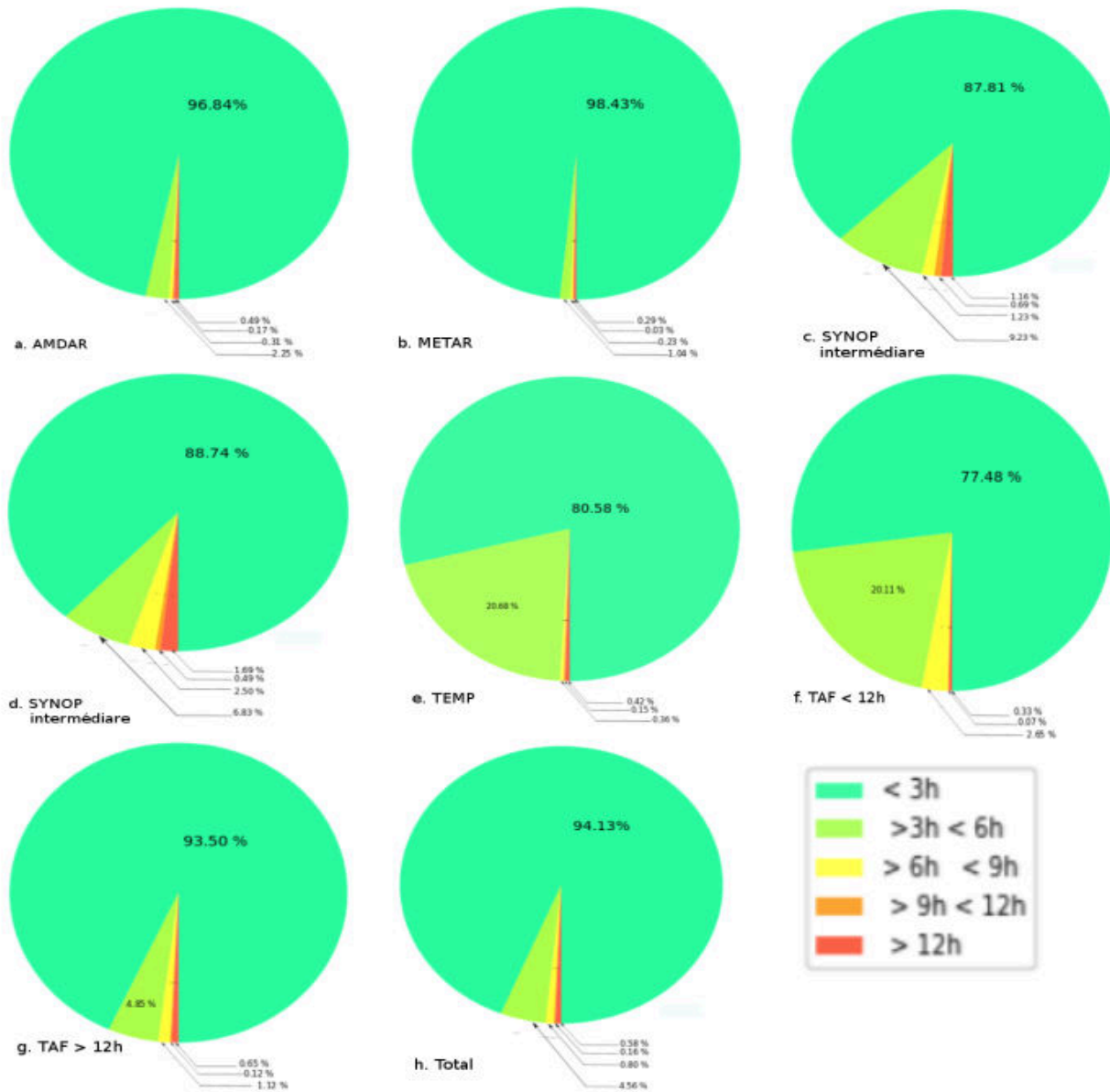


Figure 3. Pourcentage des retards dans la réception par type de message