EUMETNET
EUROPEAN METEOROLOGICAL SERVICES NETWORK

EIG EUMETNET GNSS Water Vapour Programme (E-GVAP-III)

WMO FM94 (BUFR) Specification for Ground-based GNSS Delay and Water Vapour Data

Version 1.4 – 1 December 2016

Prepared by:
Met Office

Ref: E-GVAP/METO/FMT/BUFR/001

E.I.G. EUMETNET, Avenue Circulaire 3,
1180 Bruxelles, Belgique,

Met Office

WMO FM94 (BUFR) Specification for Ground-based GNSS Delay and Water Vapour Data

Document Author Table

	Name(s)	Function	Date	Comment
Prepared by:	Dave Offiler	E-GVAP Team	1 Dec16	*Update Release*
Reviewed by:	Owen Lewis	E-GVAP WG2	1 Dec16	*Reviewed*
Approved by:	Henrik Vedel	E-GVAP Coordinator	1 Dec16	*Approved*

Document Change Record

Version	Date	By	Change
1.0	3 Dec 2009	DO	**This new E-GVAP-II document** is based on the COST-716 document *WMO FM94 (BUFR) Specification of COST-716 Processed GNSS Data,* Version 2.2 (19 Dec 2003) with the following changes: • Document under EUMETNET branding • Updated Acronyms list • Updated References list and moved to Introduction section • Added new sub-sections on BUFR Sections 1, 2 & 3 • Table 1 (draft template using local descriptors) replaced by 2 new tables showing Section1 information (Ed.3 & Ed.4) • Table 7 (was Table 6) updated to include all defined 'MetO' sub-centres • Extended section on WMO routing headers • Updated body text throughout to reflect current status for E-GVAP-II.

1.1	21 Oct 2013	DO	**Updated document for E-GVAP-III**
			• Modified title to be consistent with other E-GVAP format documents
			• Updated References and Acronyms lists
			• Split BUFR details into new Section 4 with all six BUFR Sections described
			• Inserted new tables:
			    • Table 1 (BUFR Section 0)
			    • Table 4 (BUFR Section 3)
			    • Table 5 (End of Message)
			    • Table 11 (typical message lengths)
			    • Table 13 (geographical area codes)
			    • Table 14 (summary of ARHs currently in use)
			• Added new entries to Table 9 (previously Table 7)
			• Added new 'KBOU' ICAO code to Table 12 (previously Table 8)
			• Updated body text throughout to reflect current status for E-GVAP-III, including recent NOAA GB-GNSS data on the GTS.

1.2	20 Feb 2014	DO	• Added SGOB to Table 9 • Updated Section 5 to note that all geographical area designaor (A2) codes in Table 13 may in priciple be used for GB-GNSS bulletin routing headers. This follows from a NOAA request that this code letter should better reflect the actual station coverage, notably for global networks, and subsequent implementation in the the **`cost2bufr`** encoder tool mentioned in Section 6.
1.3	20 Aug 2014	DO	• Updated [RD.6] & [RD.10] to latest versions • Updated Sections 5.3, 5.6 & 6.1 and Table 14 to include DEMO and TEST headers
1.4	1 Dec 2016	DO	• Updated links to WMO resources in Reference Decouments • Added code for Beidou to Table 7 • Added new Originating Sub-centre codes for UCAR & DITTT to Table 9 • Noted that NOAA no longer provide their data via the GTS • Updated Section 6 with software now under ROM SAF as GBGP • Updated [RD.6] & [RD.14] to latest versions • The more generic term 'GB-GNSS' replaces 'GWV' • Various minor updates to body text, acronyms list, etc.

WMO FM94 (BUFR) Specification for Ground-based GNSS Delay and Water Vapour Data

Table of Contents

List of Tables

WMO FM94 (BUFR) Specification for Ground-based GNSS Delay and Water Vapour Data

1. Introduction

EIG EUMETNET is a network grouping 26 European National Meteorological Services [RD.1], providing a framework to organise co-operative programmes between the Members in the various fields of basic meteorological activities such as observing systems, data processing, basic forecasting products, research & development and training. The GNSS Water Vapour Programme – E-GVAP (2005–2009), E-GVAP-II, 2009–2013) and E-GVAP-III (2013–2017), hereafter 'E-GVAP' – is set up to provide its EUMETNET partners with European GNSS signal delay and water vapour measurements for operational meteorology. This is being done in close collaboration with the geodetic community in Europe.

1.1 Purpose of document

This document describes the template specification which is used for the encoding of E-GVAP (and similar) processed GNSS[1] delay and water vapour[2] data (hereafter 'GB-GNSS') into the WMO FM94 (BUFR) format [RD.2]. The specification is generic in order to support the encoding of data from processing networks other than the European ones processed under E-GVAP.

BUFR formatting is a requirement for dissemination of near-real time GB-GNSS data – and most other observational data types – over the WMO Global Telecommunication System (GTS)[3], Regional Meteorological Data Communications Network (RMDCN) [RD.4] or future WMO Information System (WIS) [RD.5] – although we should note that WIS may allow the transmission of certain other formats, such as netCDF [RD.15]. BUFR is also the sole or principle storage/archive and interface format for assimilation into NWP models at many operational centres, therefore there is a requirement for BUFR, whatever the original format in which the centre may receive the data.

The BUFR template specification described in this document was submitted to WMO in 2002 and gained formal ratification for publication in BUFR Master Table Version 12 for operational use on the GTS as from November 2003. GB-GNSS data from the COST-716, TOUGH and E-GVAP projects have been disseminated on the GTS since March 2004. The current BUFR tables are at Version 25, 11 November 2015 [RD.6], but the GB-GNSS template has not changed since Version 12.

This document is intended for users wishing to encode or decode GB-GNSS data. Whilst all of the information to do this is available in the published WMO manuals and extracted tables (downloadable from the WMO website), this document provides a convenient 'one-stop-shop' to bring all the GB-GNSS-specific BUFR table details (and more) into one place.

1.2 What is BUFR?

The WMO FM94 format (otherwise known as BUFR – Binary Universal Form for data Representation) [RD.2] is a standard for encoding meteorological and other (mainly) observational data for efficient transmission and storage. The data in a BUFR 'message' are encoded into a pure binary (bit-oriented steam). The encoding and decoding process is table-driven, meaning that all data types must be pre-defined, and that these complementary processes must use the same table definitions. Once parameters are defined in this way and (eventually) accepted and published by WMO (see [RD.6]) any data within a BUFR 'message' is self-defining and can be encoded and decoded by generic routines and (with suitable wrappers for routing of WMO bulletins) be transmitted via GTS links.

---

[1] *For E-GVAP we have adopted the use the generic term 'GNSS' to encompass the US GPS, Russian GLONASS, European Galileo, Chinese Beidou and Indian IRNSS constellations even though today most receivers are only GPS-capable, though GLONASS is starting to become better supported also.*

[2] *The principle observable derived from ground-based GNSS receivers of interest to meteorology is the signal delay (expressed as an excess distance over the* in vacuuo *case). As this delay is sensitive* inter alia *to atmospheric water vapour, the term 'GPS (or GNSS) Water Vapour' or 'GWV' is in widespread common use for this data type. The abbreviation 'GB-GNSS' in this document should be taken to implicitly include delay and water vapour information derived from raw ground-based GNSS measurements.*

[3] *In this document, we will refer to the 'GTS' meaning the generic WMO-controlled networks used to exchange data between meteorological centres whatever the topology, technology or provider/owner. For our purposes, RMDCN and future WIS and similar network implementations can be considered to be part of 'The GTS'.*

Almost all operational meteorological services now transmit and receive observational data in the BUFR format and most also store (and archive) these data in BUFR, to be retrieved and decoded as needed for assimilation into NWP models etc. Indeed, several NWP centres, including the Met Office, KNMI and ECMWF can handle only BUFR in their operational assimilation systems. All new data types must therefore be encoded into BUFR before they can be used in this way. Contact with other countries known to be working with ground-based GNSS data (in Europe, via COST-716, USA, Australia, New Zealand, Japan and Taiwan) confirmed that no BUFR definition already existed, or had been formally proposed, for this data type. However, a practical – albeit local and informal – implementation had been defined for similar data processed under the EU-funded CLIMAP project [RD.7].

As part of COST-716 (and other parallel projects), the Met Office had begun running assimilation trials to assess the likely impact of GB-GNSS data in its NWP system. It followed, therefore, that these data must first be encoded into BUFR and added to the local on-line meteorological synoptic database, known as the 'MetDB'. This required a BUFR definition for the GB-GNSS data type, so a locally defined template was initially used. This later evolved into the WMO-approved template described in this document. The COST-716 BUFR specification [RD.8] documents this temporary local template, which is no longer used.

Today the BUFR template described here is used to encode all E-GVAP data centrally at the Met Office, with many of the Analysis Centres' data being disseminated on the GTS from Exeter. Until recently, NOAA/ESRL also encoded their GB-GNSS in the same BUFR template for dissemination via the GTS from Washington[4]. See Section 5 for details of the GTS routing headers.

1.3 Document outline

Section 2 of this document gives a brief overview of BUFR for readers not familiar with the system and its terminology;
Section 3 then lays out the background to how the BUFR description was initially developed for COST-716 (and later) TOUGH and E-GVAP GB-GNSS data;
Section 4 provides the detailed BUFR implementation for GB-GNSS data;
Section 5 discusses the routing headers which allow dissemination of the data via WMO communication links; and finally,
Section 6 gives the status of available software packages for encoding and decoding GB-GNSS data to/from BUFR.

1.4 Reference Documents

[RD.1] EUMETNET website: http://www.eumetnet.eu/

[RD.2] WMO (2001). *Manual on Codes*. WMO-306, Vol. 1.2, Part B, 2001 with Supplement 3, 2007. WMO, Geneva. See latest online version at http://www.wmo.int/pages/prog/www/WMOCodes.html

[RD.3] WMO (2009). *Manual on the Global Telecommunications System*. WMO-386, Vol.1, Part II, Attachment II-5 (Data Designators $T_1T_2A_1A_2ii$ in Abbreviated Headings), 2009. WMO, Geneva. ISBN: 978-92-63-10386-4. Latest (2011) version available for purchase via link at: http://www.wmo.int/e-catalog/detail_en.php?PUB_ID=498&SORT=N&q=

[RD.4] ECMWF – RMDCN webpage: http://www.ecmwf.int/services/computing/rmdcn/

[RD.5] WMO – WIS webpage: http://www.wmo.int/pages/prog/www/WIS/index_en.html

[RD.6] WMO (2015). Published tables (currently Version 25, 11 November 2015) can be downloaded from: http://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/VolumeI.2.html

[RD.7] CLIMAP (2000). *GNSS data in WMO FM94 (BUFR) format*. CLIMAP Technical report Ref. METOFFICE/CLIMAP/TN/05, May 2000

[RD.8] COST (2003). *WMO FM94 (BUFR) specification for COST-716 processed GNSS data*. Version 2.2, 19 December 2003

[RD.9] Dragosavac, M. (2009). *BUFR User's Guide*. ECMWF Operations Department Technical Note, July 2009

[4] *Having handed their processing to a commercial company, NOAA ceased GTS dissemination of their GB-GNSS data on the GTS on 3 October 2016.*

[RD.10] E-GVAP (2014). *'COST-format' file specification for ground-based GNSS delay and water vapour data*. Version 2.2a, 20 February 2014

[RD.11] WMO (2009). *Operational procedures for the GTS*. Attachment II-5 to WMO-386, 4 November 2009. http://www.wmo.int/pages/prog/www/ois/Operational_Information/Publications/WMO_386/AHLsymbols/TableDefinitions.html

[RD.12] ICAO (2013). *Demo tool for* Location Indicators: http://www.icao.int/safety/ANP/Pages/Demo-LocationIndicators.aspx
(The full and latest list of Location Indicators can be found in ICAO Publication Doc. 7910 available (at cost) via the ICAO website, publications catalogue)

[RD.13] ECMWF – BUFR software library webpage: http://www.ecmwf.int/products/data/software/bufr.html

[RD.14] ROM SAF (2014). *Radio Occultation Package – An Overview*.
Ref. SAF/ROM/METO/UG/ROPP/001. Version 8.0, 31 December 2014.
Copy available from the ROM SAF website: http://www.romsaf.org > Software > ROPP Overview Guide

[RD.15] Unidata netCDF home page: http://www.unidata.ucar.edu/software/netcdf/

Links correct at 16 Nov 2016.

1.5 Acronyms, Abbreviations & Initialisms

ARH	Abbreviated Routing Header
ASI	Agenzia Spaziale Italiana (Matera, IT)
ASCII	American Standard Code for Information Interchange
BKG	Bundesamt für Kartographie und Geodäsie (Frankfurt am Main, DE)
BUFR	Binary Universal Format for the Representation of data (WMO)
CCITT	Comité Consultatif International Téléphonique et Télégraphique
CLIMAP	Climate and Environment Monitoring with GNSS-based Atmospheric Profiling (EU)
CNRS	Centre National de la Recherche Scientifique (Nice, FR)
Compass	Chinese satellite navigation system (CN)
COST-716	Co-operation in Science and Technology Action 716 (EU)
CREX	Character form for Representation and Exchange of data (WMO)
DITTT	Direction des Infrastructures, de la Topographie et des Transports Terrestres (New Caledonia)
ECMWF	European Centre for Medium-range Weather Forecasts (Reading, GB)
E-GVAP-III	EUMETNET GNSS Water Vapour Programme (Phase 3)
EIG	Economic Interest Group
ESA	European Space Agency
ERSL	Earth System Research Laboratory (NOAA; Boulder, CO, US)
EU	European Union
EUMETNET	European Meteorological Services Network
EUMETSAT	European organisation for Meteorological Satellites
FM94	WMO Form no. 94 (i.e. BUFR)
FTP	File Transfer Protocol (under TCP/IP)
Galileo	Future European GNSS system (EU/ESA)
GB-GNSS	Ground-Based GNSS (previously 'GPSWV' or 'GWV')
GBGP	Ground-Based GNSS Package (ROM SAF)
GFZ	Helmholtz Centre Potsdam GFZ German Research Centre of Geosciences (Potsdam, DE) [formerly GeoForschungsZentrum]
GLONASS	Globalnaya Navigatsionnaya Sputnikovaya Sistema (RU)
GNSS	Global Navigation Satellite System (generic GNSS/GLONASS/Galileo/Compass/IRNSS, etc)
GOP	Geodetic Observatory Pecny (Pecny, CZ)
GPS	Global Positioning System (US)
GRAS	GNSS Receiver for Atmospheric Sounding (EUMETSAT)
GRIB	Gridded data in Binary format (WMO)

GTS	Global Telecommunications System (WMO)
IEEC	Institut d'Estudis Espacials de Catalunya (Barcelona, ES)
IESSG	Institute for Engineering Satellite Surveying and Geodesy (Nottingham, GB)
IGE	Instituto Geografico Nacional (Madrid, ES)
IGS	International GNSS Service
IRNSS	Indian Regional Navigational Satellite System (IN)
IWV	(Vertically) Integrated Water Vapour
KNMI	Koninklijk Nederlands Meteorologisch Instituut (De Bilt, NL)
LGPL	Lesser (or Library) GNU Public Licence
LPT	Bundesamt für Landestopographie (Swiss Federal Office of Topography or SwissTopo) (Wabern, CH)
MetDB	Meteorological Data Base (Met Office)
Met Office	NMS of the United Kingdom (Exeter, GB)
netCDF	Network Common Data Form (Unidata)
NGAA	Nordic GNSS Atmospheric Analysis centre (Norrköping, SE)
NKG	Nordic Commission of Geodesy (NO, SE)
NMS	National Meteorological Service
NOAA	National Oceanic and Atmospheric Administration (US)
NWP	Numerical Weather Prediction
PCD	Product Confidence Data
RMDCN	Regional Meteorological Data Communications Network (WMO)
ROB	Royal Observatory Belgium (Bruxelles, BE)
ROM	Radio Occultation Meteorology
ROPP	Radio Occultation Processing Package (ROM SAF)
SAF	Satellite Application Facility (EUMETSAT)
SGN	Service de la Géodésie et du Nivellement, Institute Géographique National (IGN) (Saint-Mandé, FR)
TCP/IP	Telecommunications Protocol / Internet Protocol
TOUGH	Targeting Optimal Use of GNSS Humidity (EU)
UCAR	University Corporation for Atmospheric Research (US)
VAR	Variational (NWP data assimilation technique)
WIS	Weather Information System (WMO)
WMO	World Meteorological Organisation
WUEL	Wroclaw University of Environmental and Life Sciences (PL)
ZTD	Zenith Total Delay (sometimes 'Total Zenith Delay')
ZWD	Zenith Wet Delay

2. BUFR – a brief overview

The reader is referred to WMO publications (notably [RD.2]) for more detailed guidance on using BUFR and for the full specification of the BUFR format and tables etc. Documentation on the latest tables is available via the WMO's web pages [RD.6]. See also the ECMWF BUFR User's Guide [RD.9].

2.1 Sections of a BUFR message

A complete BUFR message consists of six sections:

- Section 0 starts with the characters '**BUFR**' and contains the total message length and Edition number
- Section 1 is the identifier section and contains header information, type of data, date etc. The detailed content depends on the Edition.
- Section 2 is optional, and is not used in this application
- Section 3 contains one or more 'descriptors' (see later) which are entries to the look-up tables
- Section 4 contains the encoded (observation) data
- Section 5 is a terminator with the characters '**7777**'

2.2 BUFR Tables

BUFR is driven by a set of tables, known as Tables A, B, C, D and the Code/Flag Tables. Some code tables are common with other WMO forms (notably CREX) and are published as 'Common Code Tables'. The latest ratified tables are published by WMO [RD.6] and computer-compatible versions may be used in the encoding and decoding process.

- **Table A** contains a list of codes for generic data categories, such as 'satellite sounding' or 'surface observations' as coded in BUFR Section 1. Common Code Table C-13 lists data sub-categories associated with Table A entries, also coded in Section 1 (BUFR Edition 4 only). These tables are not necessary for encoding and are only used when interpreting[5] a decoded message.

- **Table B** defines data elements (parameters) such as 'temperature' by an 'element descriptor' (see below). The descriptor points to meta-data such as the parameter's name, units, a scaling factor, reference offset value and the number of bits needed to represent the valid range of scaled values for that parameter. Some elements refer to a 'code' or 'flag' table; the former is a coded integer value (e.g. satellite identifier) and the latter a related set of flag (binary) bits, such as quality information. Most meteorological variables are already defined, but some parameters, such as 'Total Delay' and GB-GNSS-specific code/flag tables did not exist so needed to be defined and added to Table B.

- **Table C** defines the meaning and usage of 'operator descriptors'. This table is not directly used by encoders or decoders and is not normally implemented for computer-compatible use, as the functions defined by Table C are implemented in the core encoding and decoding software.

- **Table D** contains 'sequence descriptors' which define short, commonly used sub-sequences (such as date and time) up to the whole 'observation' (e.g. a sounding). An entry in Table D can be defined for each new data type or the sequence can be put into Section 3 of the BUFR message. If the complete sequence is defined in Table D, then Section 3 contains just the single master sequence descriptor for that data.

- **Code/Flag** tables (and the Common Code Tables) contain the textual meaning of code values and flag bits, e.g. that a satellite classification code of 401 means GPS. As for Table A, the Code/Flag Table is not used for encoding, and is only used for interpreting a decoded message.

Both Table B and Table D are split into two areas: the 'global' set and the 'local' set. The global area is where internationally agreed and published descriptors [RD.6] are placed, and these descriptors should appear in all implementations. Global descriptors have a '*YYY*' value (see below) of less than 192. Local descriptors (*YYY* $\geq$ 192) can be defined for temporary or internal use. Since there is no central regulation for local descriptors, exchange of BUFR messages containing them could be problematic if the other centre has

[5] *The term 'interpreting' in this context means presenting decoded data in human-readable form. Interpretation is usually for diagnostic purposes and is not necessary to the machine decoding process itself.*

already defined that descriptor to be a different parameter having incompatible Table B meta-data or implies a different expanded Table D sequence. For this reason, BUFR messages using local descriptors should not be transmitted on the GTS.

2.3 BUFR descriptors

A BUFR descriptor is a three-part integer denoted by ***F XX YYY***. The first part, ***F***, defines the descriptor type and the interpretation of ***XX*** and ***YYY*** depend on ***F***:

- ***F* = 0 :** an element descriptor (an entry in Table B). ***XX*** is the element class and ***YYY*** the entry within the class.
- ***F* = 1 :** a repetition descriptor, which specifies how many times (***YYY***) the following ***XX*** descriptors are to be repeated.
- ***F* = 2 :** An operator descriptor for the temporary change of, most commonly, scale, bit width or reference value. ***XX*** identifies what meta-data to change and ***YYY*** is the modifying value.
- ***F* = 3 :** a sequence descriptor (an entry in Table D). ***XX*** is the class of descriptor and ***YYY*** the entry within the class. The expansion of the sequence is a set of descriptors, which may recursively include other sequence descriptors as long as the definition is not circular.

BUFR encoders and decoders internally expand all sequence descriptors (***F* = 3**) recursively. The resulting number of element descriptors (***F* = 0**) – allowing for repetitions (***F* = 1**) and ignoring modifying descriptors (***F* = 2**) – must be equal to the total number of data fields to be encoded or decoded.

2.4 BUFR compression

BUFR has a built-in compression mechanism that can give modest reductions in the encoded message volume (typically 2:1). This is achieved by encoding several observations together; the smallest value of each parameter (over all the observations) is first encoded, and then only the increments from this are encoded, thus requiring fewer bits than the full bit width allowed in Table B. This is automatic as far as the user is concerned and does not form part of the BUFR data (descriptor) definition. If there are significant element values which are constant (all observations from one station, for instance, where the station details are duplicated for all observations, or a parameter is unvarying – perhaps 'missing[6]) then compression for that parameter will be maximal, as the value is encoded only once and automatically replicated on decoding.

2.5 BUFR message length

Compression will be greatest when the range of values for a particular parameter is small over all the observations and smallest when the whole range of possible values is present. In general, it is not possible to predict the message length, as it will depend on the number of observations and the range of each parameter. Uncompressed messages (including single observations, which by definition cannot be compressed), however, are deterministic in their length, since all the parameter bit widths are known from Table B and can be summed together with the fixed and known ‘overheads’ of Sections 0-3 and Section 5[7].

There is no theoretical limit to the size of a BUFR message. The largest that can be accommodated in practice by the current (Edition 4) Section 0 header would be almost 17 mega-octets (megabytes) but a single bulletin of that size would be far too big for the GTS. By international agreement, as specified in the GTS Manual [RD.3], single bulletins[8] must be no longer than 500,000 octets (bytes). Prior to 5 November 2007, the limit was 15,000 octets. Since there are still legacy transmission and storage systems not yet able to handle the significantly larger limit, 25,000 octets is a good safe size to use in practice for the time being.

---

[6] *Missing or invalid data must still be encoded, as BUFR data is positional. This is accomplished by setting all the bits defined for that parameter to '1'. This implies that the maximum valid range of values of any parameter after scaling and applying the reference offset is zero to 2^{width}–2. On decoding, 'all bits set' in the BUFR message is decoded and returned to the user as missing data indicator values, usually -32768, -2147483647 or -999999, depending on the implementation.*

[7] *An uncompressed GB-GNSS message containing just one observation (ZTD sample) would be approximately 300 bytes.*

[8] *A bulletin is the BUFR (or GRIB, CREX or other WMO format) message with a wrapper or 'Abbreviated Routing Header' containing coded information on the message content for GTS routing purposes. The ARH adds some 35 bytes to the BUFR message length – see Section 5.*

For indexing efficiency in the Met Office's operational MetDB, data is preferred packaged into hourly messages, each containing data for multiple stations (and possibly for multiple times within the hour). Since the volume of data for GB-GNSS is small, we are most unlikely to reach any practical limits on message length[9]. Currently, the MetDB ingest system has an arbitrary limit of 500 observations (of any type) per message, which also imposes a practical limitation on message lengths for most types of data. The current practice is to limit single GB-GNSS BUFR messages to 500 observations (timestamped ZTD samples) and to generate multiple messages if there are more than this number to be encoded.

[9] *Tests using sample E-GVAP COST-format files indicate that for GB-GNSS data, encoded BUFR messages will be of order 8KBytes for 50 stations and 4x15-minute samples for one hour (i.e. 200 observations) using compression. Some typical actual values are shown in Table 11.*

3. GB-GNSS data in BUFR

3.1 Guidelines

The guiding principles, assumptions and objectives for developing the BUFR user requirements – and hence definitions in this document – were:

- the BUFR-encoded version of GB-GNSS data is intended for dissemination in near-real time via the GTS, and to be used operationally by NWP centres; if the BUFR is not suitable for other users/applications, these will be able to use the original COST-format[10] files [RD.10] via other dissemination methods;
- the BUFR content will be a sub-set of the data contained in the hourly (or sub-hourly) near-real time COST-format as uploaded to the E-GVAP (then COST-716) project FTP hub server;
- the majority of NWP users will apply variational assimilation methods (either full 3- or 4DVAR or 1DVAR pre-processing), using as input the zenith total delay (ZTD, and potentially gradients or slant delays) in preference to zenith wet delay, ZWD, or derived integrated water vapour, IWV. The latter two quantities – which require independent meteorological data – therefore have a lower priority for availability than the total/slant delay values;
- minimise the number new BUFR descriptors and any associated code/flag tables; re-use existing descriptors where possible, using temporary scale/bit-width changes if needed, or extend existing code tables;
- not to include static 'meta-data' parameters and flags into the BUFR, as these are of little practical use to operational NWP, and in any case can be provided off-line e.g. as a look-up table or relational database keyed by GNSS station or processing centre IDs;
- the definition should be as future-proof as possible i.e. all likely (useful) data is included now, even if initially 'missing', to avoid having to go back to WMO for modifications later;
- the definition to be based on the CLIMAP BUFR [RD.7], as implemented in the Met Office using local descriptors [RD.8] in a non-operational database (but within the operational system);
- the definition should be kept as simple as possible to minimise software development for the BUFR encoding and decoding interfaces and to minimise the complexity of any proposal to WMO and to end-users not familiar with the GNSS processing details.

3.2 Rationale

Our starting point for generating GB-GNSS data in BUFR was Version 1 of the 'COST-format' [RD.10], which was the form agreed for distribution of GB-GNSS data within the project[11] for validation and NWP trial assimilation. For widest portability, COST-format files contain data having a simple column-based plain (ASCII) text representation. Similar data from non-COST sources would only require a suitable front-end file interface for a BUFR encoder application, or such files could be first converted to COST-format prior to encoding.

Some of the parameters and individual bit flags specified in the COST-format files are not likely to be required for near-real time, operational NWP use, so not all parameters/flags were included in the BUFR specification. Omitted parameters include static meta-data such as receiver and antenna types, GNSS satellite orbit type and processing method. Omitted bit flags include how the data is sampled (mean over, or at end of, the sampling period). Such items are more efficiently collated centrally and distributed off-line, e.g. as a look-up tables or a relational database keyed by GNSS station and/or processing centre. This information is in any case required for long-term climate applications and kept in a change log.

As many existing descriptors as possible were reused in order to minimise the number of new elements needing definition in Table B. Some existing BUFR element entries (such as IWV) did not have the required precision (scaling) and/or (re-)scaled valid range (bit-width); BUFR sequences can allow for this by defining temporary changes to the Table B definitions. In some cases, an element existed which is a code table entry (e.g. a list of satellite classifiers, or scanning mode significance types), so the code/text definition only needed extending in the relevant table to include the new types.

In BUFR, as in many other areas, there is often more than one way to do something. Here, we have taken the policy of implementing the simplest, not necessarily the most efficient method. For instance, we employ a

[10] *There is a proposal to replace this inflexible column-based text-based format with a new 'SINEX-TROPO' file type having a more flexible* **`key=value`** *syntax under various section headings.*

[11] *With only minor modifications, this format continued in use for TOUGH and now E-GVAP (currently at Version 2.2).*

fixed replication count for the delays and associated parameters; in this case, delayed replication is (marginally) more efficient when much of the data in the repeat sequence is missing (as is the case for slant delays to date), but at the expense of complexity. In practice, there is little overhead in applying our approach, as the data volume is so small. Indeed, there will probably be a saving, as delayed replication implies that messages cannot be compressed unless all the observations in a message have identical replication counts.

The COST-format files [RD.10] contain fields for meteorological parameters (pressure, temperature and relative humidity); when valid, these represent *a priori* information used to derive ZWD and IWV, and are provided to allow a user to recalculate these derived parameters, perhaps using different mapping functions. The parameters – when valid – may be sourced from any one of:

- local GNSS site meteorological sensor package (automatic)
- nearby synoptic station observation (automatic or manual)
- interpolated from surrounding network of synoptic stations (automatic and/or manual)
- interpolated from NWP model fields

The first source represents new observational information that could be in principle used in NWP assimilation in its own right; however, users would have to be assured that the sensors were well-calibrated and maintained and the site exposed to WMO observing standards. The other sources are not new and may introduce unwanted correlations into the assimilation. It is therefore recommended that when meteorological data is present in the GB-GNSS BUFR[12], they are *not* assimilated but treated only as supporting data, thus obviating the necessity for any (new) code table to indicate the source of any meteorological data.

We have therefore kept the definition as simple as possible, consistent with likely operational NWP usage in the future, only defining new elements and sequences where necessary. Wider use of GNSS data in an operational environment required that these new descriptors be proposed to WMO for inclusion in the global tables for GTS dissemination. This proposal was submitted in 2002 and was approved during 2003; the updated tables (including the GB-GNSS details along with many other additions) were published by WMO (Version 12 tables) for operational use on the GTS as from 5 November 2003. GB-GNSS data were first disseminated on the GTS to NWP centres in March 2004. This template remains valid in the current Version 25 tables of November 2015 [RD.6].

[12] *At present, very few GNSS sites have a local meteorological sensor package, and few processing centres have near-real time access to synoptic or NWP data, so the majority of the BUFR messages will not have valid values for these parameters.*

4. BUFR Template Specification

The WMO Manual on Codes [RD.2] provides the formal Regulations for encoding BUFR messages. The following sub-sections give details of the various parts of a BUFR message as they are implemented for GB-GNSS data.

4.1 BUFR Section 0 (Indicator)

Section 0 provides meta-data about the message itself; it consists of a data introducer, the total length of the message and the Edition number – see Table 1.

Table 1. BUFR Section 0 (Indicator)

Octet number	Contents	Value for GB-GNSS
1–4	**BUFR** (coded in CCITT International Alphabet No. 5[13])	**BUFR**
5–7	Length of whole BUFR message (octets)	Variable
8	BUFR Edition number	3 or 4[14]

4.2 BUFR Section 1 (Identification)

BUFR Section 1 includes information on the data content (in BUFR Section 4), and so is data-type specific. Section 1 information is different between Edition 3 (as previously used to encode older GB-GNSS data) and the newer Edition 4 (as used currently).

Table 2 shows the information in Section 1 for older GB-GNSS data encoded in Edition 3, and Table 3 likewise for the current Edition 4. The Originating/Generating Centre code is taken from Common Code table C-1 (or C-11 for Edition 4) and the Originating/Generating Sub-Centre code from Common Code Table C-12. The Data Category code is from Table A and the International Data Sub-Category (Edition 4) code from Common Code Table C-13.

Some encoder/decoder implementations – such as the ECMWF library [RD.9] – use Section 1 information to generate run-time BUFR look-up table file names which depend on the local table version and the originating centre and sub-centre codes. The Section 1 header for GB-GNSS data indicates that only standard operational global table elements are used, so existing files should be loaded for this type of data whatever the centre/sub-centre code values. Recent releases (v19 or later) of the Met Office decoder use only the table version number from Section 1, but by default will use the current version of the tables.

The 'Originating/Generating Centre' coded into Section 1 would normally be that for the National Meteorological Service (NMS) encoding the data and/or injecting the data onto the GTS. For our purpose, if it is not the GB-GNSS processing (or analysis) centre doing the encoding, then the code should be that of the NMS of the country disseminating the BUFR via GTS. If the processing centre is also an NMS, the sub-centre is coded as zero. In any case, the Processing Centre identifier is encoded with the observations in Section 4. There is currently only only one Originating Centre – the Met Office (074) for all E-GVAP data with an appropriate sub-centre code for each Analysis Centre. NOAA (059) also provided data in BUFR via the GTS (with sub-centre code zero), but this has recently stopped for commercial reasons.

In the future, other countries may provide similar data – for instance, Japan, Korea, Australia and South Africa are known to be developing the near-real time local networks and might at some point provide this data in BUFR on the GTS. There is also potential for individual GB-GNSS processing (or 'analysis') centres to implement the GBGP BUFR encoder (see Section 6) and for those not also NWP centres, to send the BUFR files to their friendly local NWP centre for GTS injection (as GFZ do today, sending their Radio Occultation BUFR files generated by the ROPP encoder to DWD for the GTS).

[13] *In effect, standard ASCII characters.*

[14] *BUFR Edition 3 is now obsolete and only Edition 4 messages are permitted on the GTS since 6th November 2012. GB-GNSS data is currently encoded in Edition 4, but we include Edition 3 details here to document older BUFR messages that may be archived. Differences between Editions 3 and 4 are confined to BUFR Section 1; the GB-GNSS data itself is unaffected.*

Table 2. BUFR Section 1 (Identification) for older Edition 3 messages

Octet number	Contents	Value for GB-GNSS
1–3	Length of section	18
4	BUFR Master Table	0 (Meteorology)
5	Originating/Generating Sub-Centre (CCT C-12)	See Table 9
6	Originating/Generating Centre (CCT C-1)	074 (Met Office, Exeter)
7	Update Sequence Number	0
8	Optional Section 2 flag	0 (no Section 2 present)
9	Data Category (Table A)	0 (surface data, land)
10	Data Sub-Category (locally defined by Originating Centre)	14 (GNSS)
11	Version Number of Master Table[15]	13[16]
12	Version Number of Local Table	0 (local table not used)
13	Year of Century (2-digit) 'most typical for BUFR message content'	Year of first ZTD observation timestamp to be encoded
14–17	Month, Day, Hour & Minute 'most typical for BUFR message content'	Date/Time of first ZTD observation timestamp to be encoded
18	Pad to even number of octets	0

Table 3. BUFR Section 1 (Identification) for current Edition 4 messages

Octet number	Contents	Value for GB-GNSS
1–3	Length of Section 1	22
4	BUFR Master Table	0 (Meteorology)
5–6	Identification of Originating/Generating Centre (CCT C-1 or C-11)	059 (NOAA/ESRL, Boulder) 074 (Met Office, Exeter)
7–8	Centre Originating/Generating Sub-Centre (CCT C-12)	See Table 9
9	Update Sequence Number	0
10	Optional Section 2 flag	0 (no Section 2 present)
11	Data Category (Table A)	0 (surface data, land)
12	International Data Sub-Category (CCT C-13)	14 (ground-based GNSS)
13	Local Data Sub-Category (locally defined by Originating Centre)	14 (GNSS)
14	Version Number of Master Table	13
15	Version Number of Local Table	0 (local table not used)
16-17	Year (4-digit) 'most typical for BUFR message content'	Year of first ZTD observation timestamp to be encoded
18-22	Month, Day, Hour, Minute and Second 'most typical for BUFR message content	Date/Time of first ZTD observation timestamp to be encoded

4.3 BUFR Section 2 (Optional Data)

This section is not used for GB-GNSS, and is not present in the BUFR messages.

[15] *The Master Table Version number in which the template was first published. The encoded version number should not be changed to the current Version unless a change to Table B or Table D directly affects this data type.*
[16] *The value here was mistakenly incremented in 2007.*

4.4 BUFR Section 3 (Data Description)

This section contains the number of data sets (observations) encoded in Section 4 and flags indicating whether the data is 'observed' (*yes*) and whether they are encoded with compression or not (*yes*, unless the number of data sets is 1).

Section 3 also contains the single Table D descriptor **3 07 022**, this being the master sequence descriptor for the GB-GNSS data in Section 4. This is illustrated in Table 4

Table 4. BUFR Section 3 (Data Description)

Octet number	Contents	Value for GB-GNSS
1–3	Length of Section 3 (octets)	9 or 10[17]
4	Reserved	0
5–6	Number of datasets (observations)	1 to 500
7	Section 4 Data flags: Bit 1 = 1 for Observed data, 0 for Other data Bit 2 = 1 for Compressed, 0 for Uncompressed Bits 3–8: Reserved (set to 0)	1 (Observed, Uncompressed if a single observation) 3 (Observed, Compressed) if more than 1 observation)
8–9 **[10–]**	Descriptor *Repeat 2-octets per descriptor* Total number of descriptors = (Section length - 7) / 2	50966 (3 x 16384 + 7 x 256 + 22)
[10]	Pad byte (set to 0) – optional for Edition 4	0 (if Section length = 10)

4.5 BUFR Section 4 (Data Template)

The WMO-approved and published BUFR template sequence definition for GB-GNSS data is shown in Table 5. This is in the form of the Table D sequence of descriptors and their individual Table B specifications. Entries in **bold (blue) text** denote that a new Table B and/or Code/Flag Table entry was required, or that an existing entry in the relevant Code/Flag Table needed to be extended. These are defined in Tables 6, 7 and 8. Note that the Code/Flag Tables are not necessary for correct decoding of BUFR messages, but are used only when interpreting the decoded data.

The complete GB-GNSS BUFR master sequence of 31 element, sequence and operator descriptors is assigned to the global section of Table D as descriptor **3 07 022**. This single sequence descriptor is encoded into Section 3 of the BUFR message.

Note that within the GB-GNSS community, Field 21 is normally called 'Total Delay', Fields 168,171 are normally called 'Delay Gradient' and Field 173 is 'Zenith Wet Delay'. These element names required some brief expansion as to their physical meaning in BUFR Table B, for users not familiar with GB-GNSS data.

4.5.1 Station and processing centre identifiers

In order to be able to store otherwise duplicate station data (from different processing centres with overlapping networks) the 'Station or Site Name' parameter (data field 1 in Table 5) is coded as "***cccc-pppp***" where ***cccc*** is the IGS-style station 4-character (alpha-numeric) identifier, and ***pppp*** represents the identifier of the processing centre or a variant (see also Table 9). Examples: "**ZIMM-KNM3**", "**KIR1-GFZ**". Descriptor **0 01 015** provides for up to 20 characters, so there is scope for transparently extending this identifier string in BUFR, for instance should the need arise in the future to adopt the new-style 9-character IGS station names (***ccccnniii*** where ***nn*** is a 2-digit sub-site ID – usually **01** – and ***iii*** is the ISO-3166 3-letter country code, such as **DEU** for Germany). Obviously, applications would need to cope with such a change.

4.5.2 Zenith and slant total delays

In the replication of Total Delay and associated parameters (data fields 17–22 in Table 5), we treat Zenith Total Delay (ZTD) as a special form of the generic total (slant) delay. The sequence for ZTD will be encoded

[17] *Depends on the BUFR encoder implementation as to whether an optional Edition 4 even pad byte is inserted or not; e.g. the MetDB encoder inserts this byte and the ECMWF encoder doesn't. Edition 3 required this pad byte.*

first, but in any case will be identifiable among any valid slant delays by having the Satellite Identifier set 'missing', the Azimuth set to 0° and Elevation set to 90°.

Table 5. BUFR Section 4 (Data Template)

Data Field	Element Name	Descriptor	Table B Scale	Table B Ref.	Table B Width	Units	Comments
1	Station or Site Name	0 01 015	0	0	160	CCITT IA5	Station & Processing Centre – see Section 4.5.1
2	Year	3 01 011	0	0	12	Year	4-digit year
3	Month	"	0	0	4	Month	1–12
4	Day	"	0	0	6	Day	1–31
5	Hour	3 01 012	0	0	5	Hour	0–23
6	Minute	"	0	0	6	Minute	0-59
7	Latitude (high-accuracy)	3 01 022	5	-9000000	25	Degrees	±90º – precision equivalent to approx. 1 m
8	Longitude (high-accuracy)	"	5	-18000000	26	Degrees	±180º – precision equivalent to approx. 1 m
9	Height of Station	"	0	-400	15	Metres	to 1 m – geometric height relative to mean sea level
10	Time Significance	0 08 021	0	0	5	Code Table	23=Monitoring Period
11	Time Period	0 04 025	0	-2048	12	Minutes	5–60 mins, nominally 15 mins
12	Pressure	0 10 004	-1	0	14	Pa	to 0.1 hPa – e.g. local AWS ob.
13	Temperature	0 12 001	1	0	12	K	to 0.1 K – .e.g. local AWS ob.
14	Relative Humidity	0 13 003	0	0	7	Percent	to 1% RH – e.g. local AWS ob.
15	**Quality Flags for ground-based GNSS data**	**0 33 038**	**0**	**0**	**10**	**Flag Table**	**Bit flags – see Table 6**
16	Total Number	0 08 022	0	0	16	Numeric	No. of GNSS satellites in solution
	Replication	1 06 025					Replicate next 6 descriptors 25 times – see Section 4.5.2
17	Satellite Classification	0 02 020	0	0	9	Code Table	**GNSS series – see Table 7**
18	Platform Transmitter ID	0 01 050	0	0	17	Numeric	GNSS PRN (1-24)
19	Azimuth	0 05 021	2	0	16	Degrees	to 0.01° – clockwise from True North
20	Elevation	0 07 021	2	-9000	15	Degrees	to 0.01° – above horizontal
21	**Atmospheric Path Delay in satellite Signal**	**0 15 031**	**4**	**10000**	**15**	**Metres**	1000–4200 mm to 0.1 mm
22	**Estimated Error in Atmospheric Path Delay**	**0 15 032**	**4**	**0**	**10**	**Metres**	0–100 mm to 0.1 mm
(166)	*(end replication)*						
167	Sample Scanning Mode Significance	0 08 060	0	0	4	Code Table	**=5 (Nth/Sth) – see Table 8**
168	**Difference in Path Delays for Limb Views at Extremes of Scan**	**0 15 033**	**5**	**-10000**	**15**	**Metres**	±(0–100 mm) to 0.01 mm
169	**Estimated Error in Path Delay Difference**	**0 15 034**	**5**	**0**	**14**	**Metres**	0–100 mm to 0.01 mm
170	Sample Scanning Mode Significance	0 08 060	0	0	4	Code Table	**=6 (East/West) – see Table 8**

Data Field	Element Name	Descriptor	Table B Scale	Table B Ref.	Table B Width	Units	Comments
171	**Difference in Path Delays for Limb Views at Extremes of Scan**	**0 15 033**	**5**	**-10000**	**15**	**Metres**	±(0–100 mm) to 0.01 mm
172	**Estimated Error in Path Delay Difference**	**0 15 034**	**5**	**0**	**14**	**Metres**	0–100 mm to 0.01 mm
173	**Component of Zenith Path Delay due to Water Vapour**	**0 15 035**	**4**	**0**	**14**	**Metres**	0–1000 mm to 0.1 mm
	Change bit width	2 01 131					Set bit width to 10
	Change scale	2 02 129					Set scale to 1
174	Precipitable Water	0 13 016	0	0	7	Kg m^{-2}	0–100 Kg m^{-2} to 0.1 Kg m^{-2}
	Change scale	2 02 000					Re-set to Table B value
	Change bit width	2 01 000					Re-set to Table B value
175	Log_{10} of Integrated Electron Density	0 15 011	3	14000	13	$\log(m^{-2})$	14–22 to 0.001 – TEC=10^{value}

4.5.3 Quality Flags

The Quality Flags element for the GB-GNSS BUFR (field 15 in Table 5 and detailed in Table 6) is a combination of the COST-format's *PCDH* and *PCDD* parameters (see [RD.10]). In BUFR, combinations of bits to form a value are not allowed within a flag table, so the number of satellites is extracted to form a separate Total Number element (field 16 in Table 5). Further, PCD bits associated with the source and type of supporting meteorological data are omitted.

BUFR defines the order of bit flags as 'left to right' (i.e. the order that bits are transmitted), starting with bit 1. This convention is the opposite of that used for most computer definitions, where bit 1 (or bit 0) is 'rightmost' or least significant bit. Hence the numbering of bits in Table 6 is the opposite of that in the COST-format specification [RD.10].

Table 6. Quality Flags for GB-GNSS Data (10-bit Flag Table)

Descriptor	Bit	Meaning when clear	Meaning when set
0 33 038	1	ZTD quality is considered good	Zenith Total Delay quality is considered poor
	2	No Galileo Satellites used	Galileo satellites used
	3	No GLONASS satellites used	GLONASS satellites used
	4	No GPS satellites used	GPS satellites used
	5	No meteorological data applied	Meteorological data applied
	6	Atmospheric loading correction not applied	Atmospheric loading correction applied
	7	Ocean tide loading not applied	Ocean tide loading applied
	8	Near-real time quality data processing	Climate quality data processing
	9	Data (re-)processed off-line	Near-real time data processing
	All 10	As above	Missing

Bits for use of Beidou and IRNSS may need to be added in the future; this would require a new Table B descriptor and hence a new Table D sequence descriptor.

4.5.4 Satellite Classification

Table 7 shows additional entries in the Satellite Classification code table for the GNSS constellation (field 17 – and replications – in Table 5). A new entry for IRNSS (and/or any other new GNSS constellations) may need to be added in the future.

Table 7. Satellite Classification (Extended Code Table)

Descriptor	Value	Meaning
0 02 020	401	GPS
	402	GLONASS
	403	GALILEO
	404	BEIDOU

4.5.5 Sample Scanning Mode Significance

Table 8 shows additional entries in the Sample Scanning Significance Mode code table which prefix each of the gradient values and their errors (fields 167-169 and 170-172 in Table 5).

Table 8. Sample Scanning Mode Significance (Extended Code Table)

Descriptor	Value	Meaning
0 08 060	5	North/South
	6	East/West

4.5.6 Originating Sub-centre identifiers

Table 9 is a list of data processing centres which have at some point supplied GB-GNSS data as COST-format or equivalent files; some of these no longer supply data for E-GVAP and not all are (or have been) disseminated on the GTS.

Sub-Centre code values are encoded into BUFR Section 1, Octet 5 (Edition 3) or Octets 7–8 (Edition 4) as 'Originating/Generating Sub-Centre'.

Code Table **`0 01 034`** (Common Code Table C-12) is defined by the local Originating/Generating Centre. These particular codes are only valid when the Originating Centre code is **`74`** (Met Office, Exeter) – other sub-centre codes must be defined if another Originating Centre is encoding/disseminating the GB-GNSS data. These locally-defined sub-centre codes should be notified to WMO for inclusion in the published Common Code Table C-12.

The E-GVAP ID shown in Table 9 is the character identifier encoded as part of the Station Name field in Section 4 (field 1 in Table 5), except that some processing centres may supply alternately-processed data under a modified ID. The actual ID string is encoded into the Station Name, but the sub-centre code value in Section 1 is always that for the principle ID shown in the table. In general, such alternatively-processed data are for test purposes and are not disseminated on the GTS, though it may replace (become) the 'standard' processing at some point. Several supplier IDs are using processing systems provided and setup by GOP (CZ); being 'test' data only these are currently mapped to the sub-centre code for GOP but ought to be given their own codes if/when they become 'operational' (and thereby on the GTS).

Table 9. Originating/Generating Sub-Centres (Common Code Table C-12)

Descriptor	*Value*	*Meaning*	*E-GVAP ID*
0 01 034	0	Met Office (GB)	METO
	21	Agenzia Spaziale Italiana (IT)	ASI
	22	Centre National de la Recherche Scientifique (FR)	CNRS*
	23	GFZ, Helmholtz Centre Potsdam (DE) [formerly GeoForschungsZentrum, Potsdam]	GFZ
	24	Geodetic Observatory Pecny (CZ)	GOP
	25	Institut d'Estudis Espacials de Catalunya (ES)	IEEC*
	26	Swiss Federal Office of Topography (CH)	LPT
	27	Nordic Commission of Geodesy (NO)	NKG*
	28	Nordic Commission of Geodesy (SE)	NKGS*
	29	Institute de Geodesie National (FR)	SGN
	30	Bundesamt für Kartographie und Geodäsie (DE)	BKG
	31	Institute for Engineering Satellite Surveying and Geodesy (GB)	IES2#
	33	Koninklijk Nederlands Meteorologisch Instituut (NL)	KNMI
	34	Nordic GNSS Atmospheric Analysis centre (SE)	NGAA#
	35	Instituto Geografico Nacional (ES)	IGE
	36	Met Eireann (IE)	IRE2*
	37	Royal Observatory of Belgium (BE)	ROB
	38	Oslo GNSS Atmospheric Analysis centre (NO)	OGAA$
	39	University of Luxembourg (LU)	UL01#
	40	NOAA Earth Systems Research Laboratory (US)	NOAA*‡
	41	Wroclaw University of Environmental and Life Sciences (PL)	WUEL#
	42	Budapest University of Technology and Economics (HU)	SGOB#
	43	UCAR, SuomiNet/CONUS (US)	CONH#
	44	Direction des Infrastructures, de la Topographie et des Transports Terrestres (NC)	DITT@

Key:
** = no longer providing data.*
 # = test data for project use only – not currently disseminated on the GTS. CONH is restricted to local Met Office use only.
@ = demonstration data – disseminated on the GTS to a limited number of NWP centres (currently 1)
$ = reserved – not yet providing data.
*‡ = temporary local use – until 3 October 2016, NOAA were encoding BUFR locally and disseminating on GTS themselves with Originating Centre code **59** and Sub-centre code zero. See also Section 5.*

4.6 BUFR Section 5 (End of Message)

This section flags the end-of-message, with the delimiter shown in Table 10. If the message has been encoded correctly, the delimiter occupies the last 4 octets of the total message length indicated in Section 0. Further, the total message length should also be equal to the sum of the lengths of the individual sections (as given in Sections 1–4, plus 8 octets for Section 0 and 4 octets for Section 5).

Table 10. BUFR Section 5 (End of Message)

Octet number	Contents	Value for GB-GNSS
1–4	**7777** (coded in CCITT International Alphabet No. 5)	**7777**

4.7 Typical BUFR Message Lengths

While it would be possible to sum up the Table B bit-widths (see Table 5) and multiply by the number of observations, as noted in Section 2.5, the length of a BUFR message cannot be pre-determined when more than one observation is encoded with compression. Some typical examples of actual message lengths (without routing headers) are given in Table 11.

Note that due to end-user storage limits, we encode a no more than 500 observations per BUFR message per clock hour, so there may be more than one BUFR message generated; here we record the length of the longest message. In practice, the number of stations provided in any period will vary, and the total number of observations will also change with quality control rejections, so the message length will anyway vary, not just due to compression rates.

GTS bulletins will add 35 bytes for an abbreviated routing header and trailer (see Section 5) plus 10 bytes (the exact size depends on local GTS node requirements) for TCP/IP wrappers.

From these examples, we can safely say that with <=500 observations per message, the maximum GTS bulletin message length will be less than 20Kbytes, or an average length of around 10Kbytes.

Table 11. Typical BUFR message lengths[18]

Processing Centre ID	No. of stations	No. of ZTD samples per station	Total no. of samples (observations)[19]	No. of BUFR messages	Maximum message Length (bytes)[20]
ASI	190	4	756	2	17,232
GOPG	102	2	182	1	6,088
KNM3	50	5	250	1	7,960
LPTR	44	11	483	1	14,256
METG	99	5	477	1	16,530
METO	217	5	1070	3	16,578
NOAA	288	1	287	1	10,894
SGN	280	5	1390	3	15,280
WUEL	112	3	330	1	10,226

Note some of these IDs have test status and are not disseminated on the GTS, though all do go into the Met Office's operational MetDB

[18] *NRT COST-format files from 13 August 2013.*

[19] *Due to quality control rejections, the total number of observations may be less than the number of stations times the number of observations (ZTD samples) per station in any given COST-format file.*

[20] *Maximum bulletin length on GTS up to 4 November 2007 was 15,000 bytes - it is now 500,000 bytes so in principle, a higher limit on the number of samples per message would be possible if it were not for end-user limitations. Following a MetDB upgrade in 2012, the maximum number of observations per BUFR message was raised from 200 to 500, resulting in slightly better compression rates.*

5. GTS dissemination

A 'naked' BUFR *message* alone cannot be transmitted on the GTS; some standard headers need to be added so that messages can be routed as required.

The WMO 'Abbreviated Routing Header' (ARH) [RD.3] allows GTS nodes to route messages (e.g. data in SYNOP, BUFR or GRIB code forms) – or to accept messages for broadcast topologies – in a table-driven way without knowing the messages' data contents. The ARH is a set of characters from the International Alphabet No. 5 (CCITT-IA5 - equivalent to ASCII) and takes the form:

<SOH><CR><CR><LF>nnn<CR><CR><LF>$T_1T_2A_1A_2$ii<SP>cccc<SP>YYGGgg<CR><CR><LF>

This sequence is followed by the BUFR message ('**BUFR**'…'**7777**') and finally the end-of-message trailer sequence:

<CR><CR><LF><ETX>

The header, message and trailer are collectively known as a *bulletin*. A bulletin may contain more than one message, but typically contains only one, which will be the case for GB-GNSS data. There may be additional characters before the routing header and after the last bulletin appropriate to particular GTS transmission protocols, such as TCP/IP (FTP), the description of which is beyond the scope of this document. In any case, these 'wrapper' characters are not part of the GTS transmission and are not seen by end-users.

The elements of the WMO header & trailer are described below and summarised in Table 12, along with reserved settings for GB-GNSS bulletins. For more details of the possible values for these elements, see [RD.3].

5.1 Channel (Bulletin) Sequence Number

The **nnn** element is a sequential three-digit number in the range **001** to **999**, normally generated by the encoder or routing node. It's purpose is two-fold:

1. it allows possibly missing bulletins to be identified and re-requested
2. it allows bulletins to be re-ordered if data has to be split between bulletins and needs to be merged back (GTS does not guarantee reception in any particular order of sending).

5.2 Geographical Area Designator

When T_1 is **I** (see Table 12), the A_2 element is a Geographical Area Designator (letter code) identifying the regional location of the data in that bulletin. These location codes (from Table C3 in [RD.3]) are shown in Table 13 and can be used by routing nodes to filter bulletins by broad region without having to decode the BUFR. For GB-GNSS data, the appropriate letter is generated from the overall range of latitude and longitude values of stations encoded in each BUFR message.

For GB-GNSS data, this letter may in principle take any valid code from Table 13, though in practice the stations tend to be highly weighted to Europe, so most codes will be **A** or **D**. Some will be or **T** when the span of station locations is not contained within a single quadrant and latitude-band box (for instance *METO*). **N** is used for the Central American network of *SGNC* and **X** for global-coverage networks such as the *GOPG*, *METG* and (ex-)*NOAA* streams.

5.3 Product (Data) Identifier

The **ii** field is a 2-character product or data identifier which should make the fixed **$T_1T_2A_1A_2$ii cccc** part of the ARH unique. For operational GB-GNSS, this was chosen to be **14**, denoting GNSS data[21]. Other values may be generated for non-operational or restricted data.

Currently, data supplied in COST-format files [RD.10] can have a file status of *operational*, *demonstration* or *test*, and these are mapped to the Product Identifier value thus:

Operational: **ii=14** – for open dissemination on the GTS
Demonstration: **ii=15** – for restricted dissemination on the GTS and usually for limited time periods
Test: **ii=16** – not for dissemination on the GTS (local/internal routing only)

[21] ***ii=14*** *is also used for GNSS Radio Occultation bulletins, but with* ***$T_1T_2A_1$=IUT****. However, this value in the ARH is not reserved for, or unique to, GNSS, but it does reflects the Data Type value coded in BUFR Section 1. (The origins of this value derive from the EU COST Action 716:* ***7+1+6=14****)*

Table 12. WMO Bulletin Routing Header/Trailer elements

Field ID	Description	Value for GB-GNSS
<SOH>	'Start of Header' character: byte value = 1 decimal (01 hex)	**1**
<ETX>	'End of Transmission' character: byte value = 3 decimal (03 hex)	**3**
<CR>	'Carriage Return' 'character: byte value = 13 decimal (0D hex),	**13**
<LF>	'Carriage Return' character: byte value = 10 decimal (0A hex)	**10**
<SP>	'Space' character: byte value = 32 decimal (20 hex)	space
nnn	'Message sequence number' generated by the encoding or routing centre	**'001'–'999'**
T₁	Data Exchange format	**'I'** = Observations in binary code
T₂	Data Type	**'S'** = Surface
A₁	Data Sub-type	**'X'** = Remotely-sensed
A₂	Geographical Area Designator: **'A'–'L'**, **'N'**, **'S'**, **'T'** or **'X'** (see Table 13)	In principle, any, but normally one of: **'A'**,**'D'** or **'T'** for European data **'N'** for Northern Hemisphere data **'X'** for Global data
ii	Product type	**'14'** (GNSS)
cccc	Source of the message (ICAO Location Indicator [RD.12])	Potential codes for GB-GNSS data include (but are not limited to): **'EDZW'** Offenbach DE **'EGRR'** Exeter GB **'EHDB'** De Bilt NL **'EKCH'** Copenhagen DK **'ENMI'** Oslo NO **'ESWI'** Norrköping SE **'KBOU'** Boulder (NOAA) US **'KWBC'** Boulder (UCAR) US **'LERT'** Rota ES **'LFPW'** Toulouse FR **'LIIB'** Rome IT **'OKPR'** Prague CZ
YYGGgg	Date/time of observation where:- ***YY*** = Day of month (01-31), ***GG*** = hour (00-23), ***gg*** = minute (00-59).	Day & Time of first observation in BUFR message

Table 13. Geographical Area Designator (A_2) codes

<table>
<tr><th></th><th>0°W – 90°W</th><th>90°W – 180°W</th><th>180°E – 90°E</th><th>90°E – 0°E</th><th>45°W – 180°W</th></tr>
<tr><td>NH</td><td>A</td><td>B</td><td>C</td><td>D</td><td></td></tr>
<tr><td>Tropics[22]</td><td>E</td><td>F</td><td>G</td><td>H</td><td></td></tr>
<tr><td>SH</td><td>I</td><td>J</td><td>K</td><td>L</td><td></td></tr>
<tr><td>NH</td><td colspan="4">N</td><td></td></tr>
<tr><td>SH</td><td colspan="4">S</td><td></td></tr>
<tr><td>NH</td><td colspan="4"></td><td>T</td></tr>
<tr><td>Global</td><td colspan="4">X</td><td></td></tr>
</table>

[22] *The latitude boundaries are not specified in [RD.3]. For GB-GNSS bulletins, 'Tropics' is defined as 30°S to 30°N, NH as 30°N to 90°N and SH as 30°S to 90°S for the* **A** *to* **L** *quadrants, otherwise north or south of the Equator for codes* **N**, **S** *and* **T**.

5.4 ICAO Location Indicator

The **cccc** field is the ICAO Location Indicator [RD.12] of the originating centre. Table 12 shows the ICAO codes associated with some centres which may potentially put GB-GNSS bulletins onto the GTS (directly or indirectly). Each ICAO code has an associated BUFR Originating Centre code value which is encoded into BUFR Section 1. Note that there is no distinction in the ARH of individual Analysis or Sub-centres, only the source of the bulletin.

5.5 Date and Time

For most GTS routing nodes, the date/time indicated by **YYGGgg** must be current or the data may be blocked or rejected. 'Current' is typically defined as not more than 24 hours old and not in the future by more than 10 minutes. For GB-GNSS data, **YYYGGgg** will be the timestamp of the first ZTD sample in the bulletin (the same as that encoded into BUFR Section 1) which should normally be less than 3 hours old at the time of encoding. Data older than 24h will not be encoded.

5.6 Current routing headers

Table 14 shows the fixed part of the ARHs for those headers in past or current use for GB-GNSS bulletins (whether on the GTS or not). GTS nodes wanting to collect '**14**' headers (the others are restricted) should contact the appropriate source centre (or Exeter for European NMS users) if these headers are not already being routed to them.

Table 14. Summary of ARHs for GB-GNSS data on the GTS

AHR	*Source (Originating Centre)*	*Analysis Centre IDs*
ISX[A-L,N,S,T,X]14 EGRR[23]	Met Office, Exeter	All operational except **NOAA**
ISX[A-L,N,S,T,X]15 EGRR[24]	Met Office, Exeter	As needed
ISX[A-L,N,S,T,X]16 EGRR[25]	Met Office, Exeter	All test IDs
ISX[A-L,N,S,T,X]14 KBOU[26]	NOAA/ESRL Boulder, CO (via GTS node NOAA/NWS, Washington)	**NOAA**

[23] *Operational data, open dissemination on the GTS. Mostly* $\mathbf{A_2}$ *= A,* ***D***, ***N*** *or* ***T***

[24] *Demonstration data, dissemination on the GTS for limited periods and for selected NWP centres only.* $\mathbf{A_2}$ *as for* ***14***

[25] *Test data, not for dissemination on the GTS (local routing to Met Office operational database for passive monitoring only).* $\mathbf{A_2}$ *as for* ***14***

[26] *Operational data, open dissemination on the GTS. Mostly* $\mathbf{A_2}$ *=* ***D*** *or* ***X****. NB: these headers are no longer on the GTS*

6. Software Applications

6.1 Current status

A software application **cost2bufr**, with supporting I/O library, to encode GB-GNSS data from COST-format files to BUFR has been developed by the Met Office. The encoder application, written in Fortran 95, includes comprehensive quality control checking. WMO abbreviated routing headers (ARH) with or without TCP/IP wrappers are generated (mapped from the COST-format file status) and observations not meeting the 24-hour (or other user-defined) cut-off can be rejected. The tool encodes the BUFR to the specification described in this document, applying BUFR Edition 4.

The encoder application uses the Met Office's 'MetDB' BUFR kernel library which forms the basis for the operational MetDB database encode/decode interface. On-line Store, Retrieve and off-line Archive functions for GB-GNSS data in BUFR have been operational in the MetDB system since 20 April 2001. Data up until April 2003 was in the obsolete local table definition [RD.8] after which the current template was implemented.

A second tool **cost2cost** can be used as a COST-to-COST file 'converter' employing the same Q/C checks as the BUFR encoder. This tool can also be used to check that the file name matches the file content (see [RD.10]), and in diagnostic mode it can be used to check for incorrect formatting of the text-based data fields and detection of invalid or inconsistent data values. By comparing an input and the output files, minor format and data value anomalies can be detected. Two further tools – **noaa2cost** and **ucar2cost** – have been developed to pre-convert native NOAA or UCAR netCDF files [RD.15] to COST-format which can then be encoded to BUFR with **cost2bufr**.

The software (with supporting data, look-up tables, user documentation and build system) is now maintained by the EUMETSAT ROM SAF and is built as a portable package 'Ground-Based GNSS Package' (GBGP)[27] which is available from the ROM SAF website http://www.romsaf.org > *Software* > *GRM-92 GBGP*. The package includes user documentation in the form of a Release Notes (installation guide), a User Guide and Reference Manual (all as PDF files).

The GBGP tools are being used to encode E-GVAP data as hourly and sub-hourly NRT COST-format files uploaded to the Met Office's *FTPWEB* FTP server. The same software is also used on the operational servers which process the METO (Europe), METG (global) and METR (rapid-delivery, mostly UK) data.

6.2 Future Developments

Having been transferred to the ROM SAF under a formal Work Package, the future maintenance and support for the GBGP software is now guaranteed for at least the CDOP-3 period (2017–2022). While no additional major functionality (equivalent to ROPP's PP, FM or 1DVAR modules, for instance) is currently planned beyond the tools outlined above, developments to support, for example, modified or completely new file exchange formats, or to add new supplier (sub-centre) codes will continue to be supported. Being under the ROM SAF control (as a companion to the existing ROPP software deliverable) will ensure that formal project- and user-documentation, review, testing and release approval cycles will improve the quality, robustness and portability of the package into the future and likely well beyond the expected lifetime of the E-GVAP project. Updated versions will be posted on the ROM SAF website and registered users notified of such updates. The ROM SAF also provides a web-based Helpdesk support function.

[27] *GBGP is a re-packaged and re-branded hybrid of the previous GWV and GWVBUFR packages developed by the Met Office and provided informally on the E-GVAP hub server FTPPUBLIC; the basic code functionality is effectively unchanged. GBGP is available for no cost and to anyone, but as for any EUMETSAT SAF software package or data, potential GBGP users will need to register on the ROM SAF website and agree to a (non-open source) licence before downloading the distribution files.*