

# 气象数据格式标准化设计和实施进展

■ 王颖 薛蕾 赵芳 王旻燕

WMO进而制定并大力推进二进制的表格驱动码（TDCF），包括BUFR、GRIB和CREX码，并制定了由字符编码向表格驱动码过渡的详细计划，从2003年开始，逐步完成六大类资料由字符编码向表格驱动码的转换。我国已经完成字符编码向表格驱动码的过渡，欧洲部分国家的高空探测数据已经采用BUFR格式传输，在传输数据增加了每秒观测记录和探空气球漂移数据，实现依靠GPS高空探测技术每秒获得的探空曲线数据。

为了便于气象数据的全球交换应用，使各种气象探测资料和预报产品均以统一规定的编码格式在通信线路和网络上传输，WMO一直在推行气象数据格式的标准化。1947年国际气象组织（世界气象组织的前身）的天气委员会制定了“国际气象电码”，并在1949年1月1日开始使用，这最早的标准格式，采用传统字符编码（TAC），气象数据按照一定的格式以字符形式表示，目前仍在使用的全球地面天气报、高空综合探测和测风报告就是属于这一类编码，适应于早期通信传输电路速率低，只能以电报传输的客观条件。但字符编码格式繁多复杂的缺点，使WMO为统一格式标准，不得不根据业务发展情况不断对编码格式进行扩充和修订，日益复杂的规则也需要编解码程序随之不断扩充和完善。

随着科学技术的进步和探测技术的提高，以及计算机、通信、网络技术的发展，字符编码已无法满足需要，WMO进而制定并大力推进二进制的表格驱动码（TDCF），包括BUFR、GRIB和CREX码，并制定了由字符编码向表格驱动码过渡的详细计划，从2003年开始，逐步完成六大类资料由字符编码向表格驱动码的转换。2014年完成地面、高空等气象数据编码过渡工作。我国已经完成字符编码向表格驱动码的过渡，国际交换的数据采用TAC和BUFR格式同时传输。目前，许多国家也已完成过渡工作，如欧洲部分国家的高空探测数据已经采用BUFR格式传输，在传输数据增加了每秒观测记录和探空气球漂移数据，实现依靠GPS高空探测技术每秒获得的探空曲线数据。

针对气象观测技术发展、气象业务技术发展、科学研究发展需要，及气象参数多、数据量大等需求，国际的科研机构和IT业界也研制和推出了多种通用气象数据编码格式用于全球气象数据交换。如美国大学

大气科学协会（UCAR）提出的NetCDF网络通用数据格式和美国国家超级计算应用中心（NCSA）创建的HDF分层数据格式，目前已广泛用于气象数值模式产品和卫星数据的编码。此外，图像格式JPG和GIF、电子印刷品格式PDF、微软标准文档DOC格式、电子表格XLS格式和XML格式等格式已应用于气象服务产品的数据格式。

为加快推进气象现代化建设，建立全业务流程统一、标准的气象数据格式，2014年中国气象局部署了气象数据格式标准化工作，提出参照WMO和国际通用气象数据格式标准，确定地面、雷达、卫星、海洋、高空、气象服务产品、辐射、农业与生态、大气成分等全部气象数据的标准化格式，优化调整业务流程，解决数据格式不统一导致的上下游业务系统衔接连续性薄弱等问题，2020年基本实现我国气象数据格式的标准化的总体目标。

2016年，国家气象信息中心组织设计了气象信息化标准体系，具体规范了气象数据格式标准工作。气象数据格式标准主要包括数据文件、消息和数据流等数据对象的命名规范，以及遵循使用6种国际和行业通用标准格式（BUFR、CREX、GRIB、NetCDF、HDF、XML等格式）和部分自定义格式制定规范，进一步明确了气象数据格式标准化的技术路线。

## 1 我国气象数据格式现状

我国目前气象数据格式部分采用了国际通用数据格式，部分采用了国内自定义数据格式，主要包含WMO定义的数据编码格式、其他国际通用数据格式、国内自定义格式、面向服务的数据集数据格式等几种。数据格式存在着格式种类繁多、不标准、不统一的问题，如参加国际交换的地面、高空观测资料使用了BUFR格式和TAC格式传输到GTS，而国内又是

采用TXT格式从台站传输到国家级的，并且不同类型的地面站输出不同的TXT格式，有6种之多。又如我国新一代多普勒天气雷达探测网存在多种雷达型号（CINRAD/SA、SB、SC和CINRAD/CB、CC、CD、CCJ等），由于不同雷达型号的软、硬件设备不尽相同，雷达基数据也定义了7种格式。这些都影响、制约着气象数据综合效益的发挥，并存在缺乏与国际和业界一般采用标准格式的一致性，缺乏对气象观测、通信传输、数据加工、存储管理和应用等多个业务环节的统一考虑，数据格式的发展缺乏延续性等问题。

## 2 格式标准化技术路线

### 2.1 “参考借鉴国外、兼容适用国内”的标准格式研制方式

我国气象数据格式标准化工作，在关注和跟踪在气象部门已经采用或可能得到应用的新技术所涉及的气象数据格式国际标准和未来的发展趋势的同时，借鉴国外其他国家的标准格式经验，如WMO及国际业界推荐使用标准，并尽量保持数据编码格式本身和支持数据格式的数据表示标准的一致性。

气象标准数据格式的制定，在确保数据格式标准统一的同时，充分考虑气象服务业务的复杂性与特殊性，具备实用性与针对性。

对于国内同类型数据格式的定义，考虑未来发展的可扩展性、一致性、稳定性和延续性同样重要。

### 2.2 “数据格式一体化设计、新增数据格式准入”的业务流程和应用系统建设

根据综合统筹和配套管理原则，从“观测资料和产品生成系统—传输系统—数据加工系统—存储服务系统—应用系统”的全业务流程统筹考虑，实现数据格式一体化设计，重点解决观测和产品生成端的数据格式标准问题，及数据应用系统的标准化数据格式输入问题。

在新建气象业务系统准入规则中增加数据格式标准的验证，新建业务系统生成数据是否符合气象数据标准格式可作为业务系统验收或考核的重要依据。

### 2.3 “分批统筹推进、试点评估先行”的标准格式实施方案

在标准格式实施过程中，从新建观测和产品生成系统着手，随着观测和产品生成系统的更新换代，逐步推广到已建观测和产品生成系统。

在已建系统标准格式实施过程中，设立已有格式在业务流程和业务应用运行中的并行运行期，在一定时间内进行有效衔接，实现格式平稳过渡和切换。

各类气象数据格式标准化实施工作，均先开展试点试验工作，对应用软件和业务流程进行完善，对运行情况进行分析评估后，确定气象数据格式相关标准相对成熟后，再开展全国实施工作，积极稳妥推进业务系统调整。

## 3 格式标准化实施进展

2014年以来，我国气象数据格式标准化工作进展显著，以国内观测资料为重点，分批逐步的开展了多类气象资料格式标准化工作，近三年地面、高空、辐射、酸雨、雷达、农业气象等多种资料完成格式研制，并开展了试点试验工作，已初步形成我国观测数据格式标准体系。

在标准格式制定和推广的同时，还编写了6种国际和行业通用标准格式（BUFR、CREX、GRIB、NetCDF、HDF、XML）的应用指南，用以指导和规范后续格式标准化工作。

### 3.1 数据格式标准化进展情况

从数据格式标准化进展情况（表1）可以看出，到2020年，气象业务和科研中广泛应用的12种实时资料将全部采用标准统一的数据格式。地面、高空、海洋、辐射、大气成分、空间天气资料主要用BUFR格式表示，其中地面、高空、辐射、酸雨已完成BUFR格式模板的编制和试点运行，从2017年9月开始投入业务试运行，海洋和大气成分资料也即将在今年和明年完成格式编制。在目前业务中，天气和气候预报预测产品已经主要采用GRIB和NetCDF格式，卫星资料也已实现HDF格式的统一。雷达资料采用二进制格式，目前也已完成标准格式编制，在试点试验后，将在全国范围统一各种不同型号雷达的基数据和产品格式。农业气象、服务产品和人工影响天气资料，将统一为XML格式，目前均已开展格式编制工作。

表1 数据格式标准化进展情况表

资料类型	主要格式种类	2014	2015	2016	2017	2018	2019	2020
地面	BUFR							
高空	BUFR							
海洋	BUFR							
辐射	BUFR							
农业气象	XML							
数值预报产品	GRIB/NetCDF							
大气成分	BUFR							
雷达	二进制							
卫星	HDF							
服务产品	XML/图像							
空间天气	BUFR							
人工影响天气	XML							

■：格式编制 ■：试点试验 ■：业务运行

### 3.2 地面、高空等观测数据格式

我国地面观测数据最早使用WMO规定的TAC字符格式编报传输，在地面自动站代替人工观测后，由于TAC格式编码内容无法满足扩展后的自动站观测要素传输需要，我国自定义了TXT格式传输观测数据，但存在着传输格式多，国际交换复杂等问题。由于有不同类型自动站观测要素不同，不同时次上报观测要素不同的需求，在业务中存在多种数据格式，在国际交换时还需从TXT格式转换成TAC和BUFR格式进行交换。高空观测数据采用WMO规定的TAC格式传输标准层、特性层等不同高度层的观测数据，用TXT格式传输秒和分钟观测数据，也存在着格式不统一的问题，并且按照TAC格式规定，部分要素需降低精度编报传输，高空观测系统升级带来数据精确度增加的效益无法应用到气象现代化业务中。

地面和高空观测数据使用WMO推荐的BUFR格式可以从五方面有效的解决业务中存在的问题。

第一，BUFR格式的自描述性，可以适应不同类型自动站、不同时次传输内容不同的问题，传输内容可按需增减，地面BUFR数据格式模板内容就涵盖了地面国家站、区域站、公路交通站全部观测数据，有效的统一了地面和高空站传输数据格式。减少了台站上传文件数量，地面台站每小时上报文件数量由6个减少至2个，高空每次观测台站上报文件数据由9个减少至3个。

第二，BUFR格式的可扩展性，方便目前自动观测仪器不断投入业务使用，如雨滴谱数据，在地面BUFR格式试点期间，快速的实现了雨滴谱数据的编报和传输。

第三，国内和国际统一使用BUFR格式传输，减少了国际通信系统进行格式转换的复杂性，基于BUFR格式统一的数据表示方法，国内观测数据向国际转发时，只需按照WMO规定的格式模板对数据进行重新组织编报。

第四，BUFR格式是压缩格式，数据量小，为采用消息方式实时传输分钟数据提供了基础。在试点试验阶段，采用消息方式传输分钟和小时数据，大大提高了数据传输时效，从开始观测至传输到国家级的用时由分钟级提高到秒级。例如，地面分钟数据，从数据采集至传输到国家级用时27 s，其中数据采集和BUFR格式数据生成用时25 s，省级快速质控用时4 ms，传输用时小于2 s。地面小时数据，从数据采集至传输到国家级用时70 s，其中数据采集和BUFR格式数据生成用时68 s，省级快速质控用时7 ms，传输用

时小于2 s。有96%的台站地面小时数据在整点后90 s到达国家级，与现行业务中新Z文件10 min可到达99%比较，大大提高了时效。

第五，BUFR格式是二进制压缩格式，在不增加传输数据量的基础上，最大限度的保存了气象要素的原始观测精度，不会受传输数据量的限制在传输格式中减少数据精度。如高空观测数据，原TAC格式每个数据组均为5位，如果5位不够，就会降低数据精度编报。再比如位势高度，在BUFR格式中精度为1位势米，TAC格式中，500 hPa以下为位势米，500 hPa以上位势高度达到了6位，无法表示，单位改为了位势什米。温度数据，BUFR格式中是实际观测值，TAC格式报文中最末一位数字即表示小数点后1位的数值，也表示温度的正负，0和偶数为正，奇数为负，所以如果观测温度值为10.5 °C，只能编报为104，解码后即10.4 °C。风向数据，BUFR格式风向精确到度，TAC格式中风向只精确到10°。

2017年9月开始，地面、高空、辐射、酸雨数据BUFR格式在全国国家级台站投入业务试运行，标志着我国观测数据格式和数据表示方法正在逐步与国际接轨。

### 3.3 雷达资料格式

一直以来，由于新一代天气雷达型号、软硬件和生产厂家不同，全国范围雷达基数据格式有7种之多，下游应用系统不得不开发多个读取接口，以适应格式的多样性。

按照国际业界惯例，统一后的雷达基数据仍采用二进制格式，但内容涵盖了原7种格式雷达基数据的全部内容，并考虑了新增的双偏振雷达观测数据需求，对基数据格式进行了扩展，以彻底解决我国多年存在的雷达基数据格式不统一的问题。

2016年雷达基数据标准格式在19个雷达站开展了试点工作，试点范围涵盖了SA、SC和CD三种新一代天气雷达型号和新型双偏振雷达，2017年将试点范围扩展到54个雷达站，并采用数据流方式进行传输试验。试验结果表明，全体扫雷达基数据从雷达体扫开始至国家级流传输服务端收到并生成基数据文件为止，平均约343 s内到报率达到90%；现行业务的雷达基数据文件，从雷达体扫开始至国家级通信业务系统CTS1收到基数据文件为止，平均约442 s内到报率达到90%，标准格式雷达基数据流时效提高了100 s。

安徽省短临预报系统（SWAN）试验并实现了标准格式雷达基数据流的应用对接，通过虚拟体扫客户端仿照实时业务调取和显示雷达基数据，与现行业务

虚拟体扫对比, 时效提高约3.83 min, 时效改进效果明显, 并且虚拟体扫和常规体扫拼图剖面无明显差异。

#### 4 数据格式标准化发展展望

完善的气象数据格式标准体系是气象信息化发展的基石, 虽然气象数据格式标准化已经取得了一定的进展, 但距离2020年基本实现我国气象数据格式标准化的总体目标还相距甚远。部分资料类型标准格式还未制订或未完成, 如农业气象、大气成分、空间天气等。在标准格式制定的同时, 标准格式试点试验和评估、格式转换应用工具、标准格式业务准入制度等相关工作也至关重要。

前期的格式标准化工作重点是气象观测数据格式的统一, 但在气象信息化业务中, 数据存储和归档、数据产品交换、数据服务过程中的格式标准也必不可少, 在推广观测数据标准格式的同时, 结合气象大数据云平台建设, 逐步推进数据产品、数据存储、数据服务的标准格式。

在标准格式实施过程中, 试点试验是标准格式数据平稳无缝隙代替旧格式, 切换到业务系统的关键。试点试验之后的评估分析, 是优化应用程序和业务流程的参考依据, 评估分析不仅包括试点阶段业务运行和数据完整性及正确性, 应加强标准格式气象数据对业务应用和影响的分析, 首先要组织业务系统对接标准格式数据, 分析标准格式数据对业务应用的影响, 和分析标准格式气象数据与旧格式气象数据是否存在系统偏差和非均一问题, 为气象数据后续使用提供支撑。如天气雷达基数据标准格式, 在试点试验阶段测试了业务系统对接情况, 对业务用户反映的雷达类型描述不全、数据类型掩码定义不够清晰、缺少径向数据头数据类型的问题进行了优化。在后续雷达基数据标准格式推广工作中也考虑了业务系统对接过程中提出的标准格式与旧格式变动较大, 需要对原解码程序做较大改动, 业务系统的加载数据流程也需要相应调整, 根据具体情况延长标准格式和旧格式的并行期, 在全国稳步分批推广标准格式。

开发和推广标准格式相关应用程序和格式转换工具, 为标准格式应用提供方便良好的技术支持, 也是标准格式能够推广和应用的必要条件。如BUFR编码格式, 虽然WMO已经制定了技术规范并一直致力于推广应用, 但国内对BUFR格式数据处理和应用技术还不够成熟, 需要标准格式编解码软件、数据显示工具、格式转换工具支持BUFR格式数据的广泛应用。

气象数据格式标准化的软件或技术装备的市场准入, 要通过法定的方式确定, 应结合技术装备许可管

理办法等修订工作, 除针对观测系统的硬装备, 补充数据标准化格式和软件的要求, 将气象数据格式标准纳入到管理办法, 作为软件或技术装备市场准入的条件之一。

随着气象监测、预报预测、公共服务业务的发展, 气象大数据时代已悄然来临, 体量巨大、内容丰富的气象数据如何有效管理并应用是摆在气象信息技术人员面前的课题, 气象数据格式标准化工作也必须从制定和推广标准格式, 向规范标准格式制定方法, 规范数据表示方法过渡, 只有让数据格式标准化技术和方法深入人心, 在气象信息化业务建设、管理和应用工作中自觉遵循, 才能使数据格式标准化工作进入良性循环, 支撑气象大数据高效管理和应用。

#### 深入阅读

国家气象信息中心通信台, 2005. 表格驱动码编码手册: BUFR、GRIB和CREX编码. 北京: 气象出版社.  
赵芳, 2007. 气象代码的应用现状及向表格驱动代码过渡的影响分析. 应用气象学报, 18(5): 709-715.

(作者单位: 国家气象信息中心)

### 未来五年气象信息化发展主要指标

指标	2015年	2022年
1. 气象观测资料总体可用率 (%)	90	98
2. 观测资料质量控制覆盖率 (%)	70	100
3. 跨部门共享数据流量 (GB/日)	600	1000
4. 数据存储能力 (PB)	<20	100
5. 气象云计算能力 (万核)	-	9
6. 高性能计算机系统运算能力 (PFlops)	<2	20
7. 国家级主中心与异地备份中心间接入线路带宽 (Mbps)	-	2000
8. 省级中心接入线路带宽 (Mbps)	36	200
9. 气象数据服务网站访问能力——并发数	-	10000
10. 气象数据服务网站访问能力——访问频次 (亿次/日)	-	10
11. 网格气象预报产品时间分辨率 (小时)	3	1
12. 网格气象预报产品空间分辨率 (公里)	10	1
13. 高时空分辨率模式生成后到达预报员桌面时间 (分钟)	60	10

——摘自:《气象信息化发展规划(2018-2022年)》, 2017年12月