

肖华东,孙婧,孙朝阳,等. 中国气象局 S2S 数据归档中心设计及关键技术. 应用气象学报,2017,28(5):632-640.
DOI: 10.11898/1001-7313.20170511

中国气象局 S2S 数据归档中心设计及关键技术

肖华东^{1)2)3)*} 孙 婧¹⁾ 孙朝阳⁴⁾ 聂元丁¹⁾ 赵春燕¹⁾
郭 锋¹⁾ 常 飏¹⁾ 张新诺¹⁾ 刘立明¹⁾

¹⁾(国家气象信息中心,北京 100081)

²⁾(中国科学院测量与地球物理研究所,武汉 430077)

³⁾(中国科学院大学,北京 100049) ⁴⁾(国家气候中心,北京 100081)

摘 要

中国气象局 S2S(Sub-seasonal to Seasonal)数据归档中心建设是中国气象局承担世界气象组织(WMO)的世界天气研究计划(WWRP)和世界气候研究计划(WCRP)任务,由国家气象信息中心负责设计和实现。该文介绍了 S2S 数据归档中心建设中涉及的数据交换、数据检查及处理、数据归档存储及数据服务门户全流程系统设计和实现。针对各业务中心生产的 S2S 数据配置差异较大造成数据交换、同步较难这一问题,采用基于 FTP(file transfer protocol)的数据推送和基于 ECMWF(European Center for Medium-range Weather Forecasts)WebAPI 主动数据下载相结合的方式,说明数据交换、同步方法和策略。由于 S2S 数据量巨大难以高效管理服务,已设计统一的数据组织形式和存储规则,实现根据数据检索条件解析获取数据存储位置,提供便捷的数据检索下载服务。自 2015 年 11 月 15 日中国气象局 S2S 数据门户系统对外开放,目前数据门户系统已有超过 18 个国家的 300 个用户注册并下载数据。

关键词: S2S; 集合预报; 回算预报; 动态配置

引 言

业务预报和专业预报对中期预报(预报时效长达 15 d)与长期或季节预报(预报时效长至 3~6 个月)尺度之间的预报需求增长迅速^[1]。次季节-季节预报研究项目(s2sprediction.net)由世界天气研究计划/世界气候研究计划联合建立,重点研究两周至两个月时间尺度的预报,目的在于改善次季节至季节尺度预报技巧。该项目主要交付物是大量的数据集,包括从 11 个业务气象(气候)中心生产的**次季节预报(预报时效长达 60 d)**,包括延后 3 周的近实时集合预报和回算集合预报,其中部分中心来自 TIGGE(THORPEX Interactive Grand Global Ensemble)数据集生产者。预报技巧高的次季节-季节

预测(**预报时效超过 2 周,少于 1 个季度**)可提前通知管理决策者应对极端事件的风险处理和资源管理的优化决策。尽管次季节-季节预报还存在很多挑战,但在天气、气候科学及模式发展领域的投资可预期获得回报^[2-3]。

S2S 数据集自 2015 年 5 月向研究人员开放,应用该数据集可促进对次季节-季节尺度预报机理的理解,而次季节-季节预报在很长一段时间内被看作是预报的“沙漠地带”。特别是 S2S 数据集将有助于确定次季节-季节预测模式模拟和预报可预测性来源问题^[2-4]。S2S 数据集建立在 TIGGE 数据集建设的经验之上,可视作 TIGGE 数据集向着更长预报时效的延伸^[2]。S2S 数据集与 TIGGE 数据集的关键区别在于 S2S 数据集具有回算集合预报数据,而 TIGGE 数据集没有^[2,5-6]。对于短期天气预报,

2017-02-04 收到,2017-08-14 收到再改稿。

资助项目: 国家自然科学基金项目(41674085,41604009)

* email: xiaohd@cma.gov.cn

模式误差通常不是影响预报水平的控制因素,故回算预报数据集不必提供;但对于次季节-季节尺度的预报,模式误差不能忽略。因此,需要一套跨越数年的回算预报数据集以计算模式偏差。这类回算预报数据在某些情况下可用于评估预报技巧。S2S 中的模式一般也不同于 TIGGE 模式。如 S2S 模式可能是大气分量与海洋模式和海冰模式耦合的模式。

中国气象局此前承担了 TIGGE 数据归档中心建设任务,此次也应邀承担 S2S 数据归档中心建设任务^[6-7]。作为 S2S 数据归档中心,一方面可增强中国气象局对世界气象组织(WMO)的贡献,利于提高中国气象局在世界范围内的影响,另一方面也利于国内用户更为方便广泛地获取各中心的 S2S 数据。中国气象局 S2S 数据归档中心作为两大 S2S 数据归档中心之一,与中国气象局的其它信息共享系统或数据存储管理管理系统不同^[8-13],S2S 数据归档中心建设只针对各中心的 S2S 模式数据,专门面向国内外科研应用。S2S 数据归档中心要求建立用户界面友好的数据集,并提供数据集的开放访问,国家气象信息中心于 2015 年启动了中国气象局 S2S 数据归档中心建设,包括数据交换、数据质量检查、存储归档和数据服务等功能,实现对 S2S 数据集的完整正确归档和开放服务。

1 S2S 数据特征及归档中心建设需求

1.1 S2S 数据特征

S2S 数据集包括近实时集合预报和回算集合预报数据,长达 60 d 预测,来自 11 个中心,分别是澳大利亚气象局(BoM)、中国气象局(CMA)、欧洲中期天气预报中心(ECMWF)、加拿大环境及气候变化部(ECCC)、意大利大气科学和气候研究所(ISAC)、俄罗斯水文气象中心(HMCR)、日本气象厅(JMA)、韩国气象局(KMA)、法国国家气象研究中心(CNRM)、美国国家环境预测中心(NCEP)、英国气象局(UKMO)。S2S 数据均采用 GRIB2 格式编码^[14],产品水平分辨率均为 $1.5^{\circ}\times 1.5^{\circ}$ 。

由于 S2S 预报产品并未专门针对 S2S 商定的协议而生产,近实时预报产品规定只能在产品发布 3 周之后可用。表 1 说明了 S2S 数据集配置的主要特征。S2S 存储归档中要求的气象要素个数多达 46 个,包括多个气压层次的标准气象要素,还有大量的单层气象要素包括热力要素、水动力要素、地表通量要素等。但有些模式只提供要求气象要素的一部分。每个模式提供的气象要素列表可从 s2s.cma.cn/description 获知。

表 1 S2S 数据生产中心及其模式运行系统配置
Table 1 Centers that provide data to the S2S project together with the configuration of their systems

模式	预报时长/d	分辨率	近实时集合预报		回算集合预报			
			成员	频次	类型	预报时段	成员	频次
BoM	0~60	T47L17	33	2 次/周	固定	1981—2013 年	33	6 次/月
CMA	0~60	T106L40	4	1 次/d	固定	1994—2014 年	4	1 次/d
ECCC	0~35	0.6°×0.6°L40	21	1 次/周	动态	1995—2014 年	4	1 次/周
ECMWF	0~46	T639-319L91	51	2 次/周	动态	过去 20 年	11	2 次/周
HMCR	0~63	1.1°×1.4°L28	20	1 次/周	动态	1985—2010 年	10	1 次/周
ISAC	0~32	0.75°×0.56°L54	40	1 次/周	固定	1981—2010 年	1	6 次/月
JMA	0~34	T319L60	25	2 次/周	固定	1981—2010 年	5	3 次/月
KMA	0~60	N216L85	4	1 次/d	动态	1996—2009 年	3	4 次/月
CNRM	0~61	T255L91	51	1 次/月	固定	1993—2014 年	15	2 次/月
NCEP	0~44	T126L64	16	1 次/d	固定	1999—2010 年	4	1 次/d
UKMO	0~60	N216L85	4	1 次/d	动态	1996—2009 年	3	4 次/月

表 1 中,类型是相对回算集合预报而言,分为静态配置预报和动态配置预报,静态配置预报指模式版本固定,若干年的回算集合预报都是基于固定版本的模式产生,只产生一个循环的若干年的回算预报数据;而动态配置预报指回算预报和近实时预报在同一月份和日期启动,区别在于产生若干年中同一日期的数据,每完成一年的实时预报就会产生一

个循环多年的回算预报。
由表 1 可知,不同中心近实时预报的主要差别在于预报时长范围从 32 d 到 60 d 不等;大气模式水平和垂直分辨率从 30 km 至几百公里不等;集成员数从 4 个到 51 个不等;预报频次也不同。有些模式具有很大的集成员数,1 周运行 2 次(如 ECMWF,BoM,ECCC),而另一些模式具有较小的集合

成员数,每天连续运行(如 NCEP, UKMO, CMA, KMA)等。其他模式(如 CNRM)则 1 个月运行 1 次。

不同中心模式的回算预报系统配置也不尽相同。一些模式回算预报产品覆盖超过了 30 年(如 JMA, BoM),另一些则覆盖较短的年限(如 NCEP, UKMO)。一些回算预报产品以循环方式产生,而其他的则是先于业务实现前一次产生(如 NCEP, CMA)。集合成员数从 1 个(如 ISAC)到 33 个不等(BoM)。一些模式每天产生回算预报产品(如 NCEP, CMA),而有些则每周产生 2 次回算预报产品(如 ECMWF, HMCR),其他的每个月产生一次回算预报产品。

当前的 S2S 数据定时更新 11 个数据生产中心的近实时数据和动态配置的部分中心回算数据。S2S 数据属于集合预报,包括近实时集合预报(实时预报)数据和回算集合预报(回算预报)数据。每个中心各自的实时预报和回算预报时长一致,采用的模式也相同,只是在集合成员数量及生产频次上有差别。各个中心之间的实时和回算预报差异较大。另外,数据组织呈现形式复杂,各中心也大为不同,生产频率也不一致。S2S 数据生产频次复杂,每批次成员多,预报时效长,回算预报年份长久,直接体现数据量大的特征,同时意味着收集处理过程复杂。

1.2 归档中心建设需求

S2S 数据归档中心建设主要任务是收集各 S2S 数据生产中心提供的数据,并予以长期保存,同时对国内外用户提供数据检索下载服务。中国气象局国

家气候中心作为 S2S 数据生产中心的一员,向另一 S2S 数据归档中心(ECMWF)提供该中心生产的 S2S 数据,也是中国气象局 S2S 数据归档中心建设的一个任务。

S2S 数据归档中心建设需求涉及到数据交换、数据质量控制及检查、归档存储、数据服务 4 个方面。数据交换包括将中国气象局国家气候中心生产的 S2S 数据提供给 ECMWF 和收集其他 S2S 数据生产中心提供的数据;作为 S2S 数据生产中心,需要及时高效地将中国气象局国家气候中心生产的 S2S 数据提供给 ECMWF,保证完整性和正确性;作为 S2S 数据归档中心,需要完整收集各 S2S 数据生产中心提供的数据,并进行数据质量检查,发现问题时记录并通知另一归档中心 ECMWF 及数据生产中心;经过必要的加工处理后归档存储;建立友好开放的 S2S 数据门户系统,供国内外用户检索下载数据。

S2S 日常数据收集具体信息见表 2,其中回算预报接收的数据提供多个年份同月份同日的数据,故接收起始日期未标注年份。

根据 S2S 数据集配置及已有 S2S 数据收集情况,在各中心 S2S 数据生产模式配置不变的条件下,S2S 数据年增量约 10.3 TB,具体数据增量见表 3。除 ECMWF, ECCC, HMCR, KMA, UKMO 每个循环年产生 1 次多年的回算数据外,其他 S2S 数据生产中心只产生 1 次多年的回算数据。除去目前尚未接收 ISAC 和 HMCR 的回算数据外,固定配置的回算预报数据量达 25.5 TB。

表 2 S2S 数据收集情况
Table 2 Status of S2S data acquisition

中心	实时预报				回算预报				
	每次数据量/GB	产生频次	接收起始日期	数据量/GB	每次数据量/GB	产生频次	预报时长/a	接收起始日期	接收数据量/GB
CMA	1.2	1 次/d	2015-01-01	718	1.2	1 次/d	21	01-01	9198
ECMWF	8	2 次/周	2015-03-05	1236	16	2 次/周	20	04-23	
	12	2 次/周	2015-05-14		50	2 次/周		05-25	4575
NCEP	2.7	1 次/d	2015-04-29	954	7.8	1 次/d	12	03-31	2800
JMA	3.3	2 次/周	2015-06-09	224		3 次/月	32	01-01	900
BoM	3.6	2 次/周	2015-04-19	362	2.75	6 次/月	33	01-01	7400
CNRM	22	1 次/月	2015-04-01	405		2 次/月	22	01-01	3300
KMA	0.89	1 次/d	2016-01-05	106	9.66	4 次/月	14	01-01	114
HMCR	5.3	1 次/周	2015-11-25	151		1 次/周	26		
ECCC	2.7	1 次/周	2016-07-07	16	10.2	1 次/周	20	07-07	60
UKMO	1.6	1 次/d	2015-12-01	316	27.6	4 次/月	23	03-01	687
ISAC	6.6	1 次/周	2015-11-23	241		1 次/5 d	32		

表 3 各中心 S2S 数据年增量
Table 3 Increasing volume of S2S data by year

中心	实时预报/GB	回算预报/GB
CMA	445	
ECMWF	1248	5200
NCEP	986	
JMA	343	
BoM	374	
CNRM	264	
KMA	325	
HMCR	276	
ECCC	140	
UKMO	584	
ISAC	343	

2 S2S 数据归档中心设计

中国气象局 S2S 数据归档中心设计以 S2S 数据为中心,依据数据流向,设计数据交换子系统、质量控制数据处理子系统、存储管理子系统、数据服务门户。

2.1 功能设计

数据交换子系统是 S2S 数据归档中心建设的重要环节,包括将中国气象局国家气候中心生产的 S2S 数据向另一归档中心 ECMWF 发送和收集其他 S2S 数据生产中心提供的数据。建立 FTP(file transport protocol)数据下载服务,将中国气象局国家气候中心生产的 S2S 数据传输至位于 DMZ(demilitarized zone)区域的 FTP 服务器上,供 ECMWF 下载获取。收集各中心生产的 S2S 数据采用从 ECMWF 获取的方式,无需直接与其他 S2S 数据生产中心联系。数据收集必须保证数据传输的完整性和正确性,首先是建立基于 FTP 服务的数据传输方式,其次提供数据接收后的临时存储空间。考虑到

与另一数据归档中心 ECMWF 之间的数据同步,收集中涉及到日常数据获取、缺失数据补充获取、遗漏传输的静态配置回算集合预报数据获取、日常数据获取中断弥补这几种情况的处理。

数据质量控制检查子系统是在数据传输完整和正确的前提下,对打包文件中的具体 GRIB2 数据进行检查,主要检查 GRIB2 数据文件大小是否非零、格式是否正确、GRIB2 数据中气压要素场数与参考数量是否一致(进一步需检查每个气象要素场是否与参考值对应)、GRIB2 数据关键属性是否正确以及位图编码检查等。如果检查不正确经过确认后通知 ECMWF 处理。数据预处理主要是在数据质量控制检查中涉及到需要对不同组织形式的数据进行统一操作。

数据归档存储子系统主要是对数据进行归档存储,设计具有自描述信息的文件命名方式,结合在线文件系统进行存储管理,实现近实时和历史数据的存储管理,同时保证数据高效检索访问。

数据服务门户子系统包括用户管理模块、通知内容管理模块、模式数据说明模块、数据检索下载模块、数据下载统计模块,主要是建立数据服务门户,对国内外用户提供友好高效的数据检索下载功能。

2.2 运行流程

中国气象局 S2S 数据归档中心建设的主要任务是实现 S2S 数据的收集、质量控制及处理、归档存储及服务,业务逻辑具备流程化作业的特点,核心业务单元之间基于预定义的流程进行任务协作和数据交互。根据业务处理逻辑,可将业务处理过程划分为 4 个阶段:数据收集阶段、质量控制及处理阶段、数据归档存储阶段、数据服务阶段。具体信息流程图见图 1。

数据收集子系统主要是通过 FTP 方式及基于

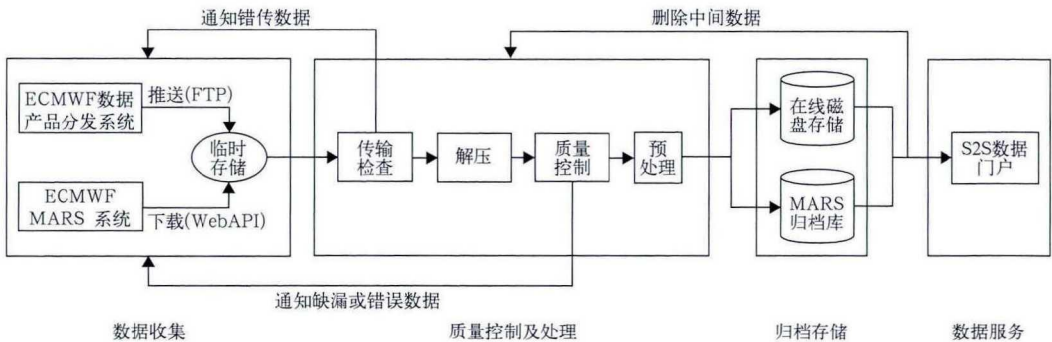


图 1 S2S 数据收集归档服务流程

Fig. 1 The flow of S2S project data collecting, archiving, and service

ECMWF 的 WebAPI 的下载程序收集 S2S 数据并置于临时存储^[15]。质量控制及处理子系统中传输检查部分记录传输正确文件列表信息,若有必要通知错传数据信息,收到传输正确的文件列表信息后,对打包压缩的数据文件解压,并予以质量控制。对检查正确文件进行预处理,若有必要则通知遗漏或错误数据信息。归档存储管理子系统将经过预处理的数据一方面传输至在线磁盘存储,一方面归档进入 MARS (Meteorological Archival and Retrieval System) 系统^[16-17]。最后数据服务门户响应用户数据检索请求,从归档存储子系统中提取数据返回下载链接给用户,实现数据服务。数据收集系统、质量检查处理子系统、归档存储子系统将各自的监控信息的记录提供给流程控制系统,用于系统监控管理。

中国气象局国家气候中心生产的 S2S 数据从内网上传至 DMZ 区域的一台 FTP 服务器,供 ECMWF 直接下载。其归档处理流程则直接在高性能计算机上经过质量检查处理后直接传输至在线磁盘存储并归档入库 MARS 系统。

3 归档中心实现关键技术

3.1 数据同步

中国气象局 S2S 数据归档中心和 ECMWF S2S 数据归档中心之间的数据同步采用两种方式:基于 FTP 服务的数据推送方式和基于 ECMWF WebAPI 主动数据下载方式。其中,基于 ECMWF 数据推送方式针对两类数据,一类是各 S2S 数据生产中心在其固定数据生产日期收集后的数据,另一类是遗漏传输或传输出现错误需要重传的数据。第 1 类数据组织形式为各 S2S 数据生产中心提供的原始形式,第 2 类数据一般先由中国气象局国家气象信息中心提供需要重传的数据列表,再经由 ECMWF 工程师直接从其 MARS 磁带库中复制至其本地磁盘的多个容量巨大的 GRIB2 文件。基于 ECMWF WebAPI 主动数据下载方式主要是针对部分 S2S 数据生产中心以磁盘或磁带等介质方式一次性提供的回算集合预报数据,亦可作为需要重传的数据获取的补充。

对于日常的数据接收,为改进数据接收的时效,自 2015 年 12 月 1 日起,由过去中国气象局国家气象信息中心定时从 ECMWF 下载其收集的其他 S2S 数据生产中心的数据方式,调整为由 ECMWF

自收集到其他 S2S 数据生产中心提供的数据即刻推送至中国气象局的数据收集 FTP 服务器指定目录下。对于缺失数据补充获取,通过设定程序定期检查缺失数据日期,提供列表清单,通过与 ECMWF 工程师协作在其 MARS 服务端命令行下直接从磁带库中复制数据,并推送至中国气象局的数据收集服务器指定目录下。对于部分中心一次性提供的回算集合预报数据获取,主要基于 ECMWF 的 WebAPI 编写 Python 程序检索获取。

3.2 数据质量检查及处理

传输中数据质量检查主要检查数据文件个数及大小、tar 打包文件及 GRIB2 格式是否正确。根据命令语句 `tar tvf *.tar` 及 `grib_get-7 *.grib` 初步测试打包文件及 GRIB2 数据文件的正确性。

数据质量控制检查主要是对解压后的 GRIB2 文件大小、GRIB2 格式、GRIB2 数据文件气象要素场数等进行初步判断,然后比照各气象要素场参考定义,对 GRIB 数据中气象要素场属性进行检查,如气象要素科目领域代码 (discipline)、气象要素类别代码 (parameter category)、气象要素代码号 (parameter number) 以及层次信息如起始固定地面类型 (type of first fixed surface)、起始固定地面缩放因子 (scale factor of first fixed surface)、起始固定地面层缩放值 (Scaled value of first fixed surface)、物理变量是否和定义一致,此外,检查 GRIB2 数据属性 GRIB 主表版本号 (gribMasterTablesVersionNumber)、集合成员数 (Number)、参考时间取值 (significanceOfReferenceTime) 及经纬度范围是否正确等^[18]。

在 GRIB_API 提供的 `tigge_check` 程序基础上,扩充几个气象要素场的检查功能,并增加数据统计类型 (瞬时值、平均值、累积值) 检查,检查 GRIB2 属性。

数据处理主要是将各个 S2S 数据生产中心提供的数据组织成统一的形式,转换成若干个单中心、单日期、单气象要素场的数据文件。为区分近实时集合预报数据和回算集合预报数据,在文件名中增加模式循环年字段,形式为 `s2s_[centre]_${cyear}_[dataDate]_[shortName].grib`。其中, `centre`, `dataDate` 和 `shortName` 分别为中心代码、数据日期、气象要素场缩写,取值从 GRIB2 文件数据属性中获取; `cyear` 则表示数据生产的年份,用以区分近实时集合预报数据、固定配置回算集合预报数据、动态配置回算集合预报数据。对于实时集合预报数

据,cyear 取值为 GRIB 数据文件中 year 属性值;对于固定配置回算集合预报数据,cyear 取值为数据生产的年份(除 ECCC 为 2016 外,其余均为 2015),动态配置回算集合预报数据,cyear 取值是从 GRIB2 数据文件 YearOfModelVersion 属性中提取的年份。当 cyear 和 dataDate 中提取的年份相等时,即为近实时集合预报数据,固定配置的回算集合预报数据中 cyear 取值不变,动态配置的回算集合预报数据 cyear 取值区间随着数据生产年份变化。

日常的推送数据、基于 ECMWF MARS 直接从磁带库中复制后推送的数据以及基于 ECMWF WebAPI 下载的数据,是以不同组织形式的 GRIB2 文件呈现,针对每个中心每个批次数据,进行排序合并处理形成一个超大的 GRIB2 数据文件,然后利用 GRIB_API 软件的 grib_filter 命令应用规则“write s2s_[centre]_ \${cyear} _[dataDate] _[shortName]. grib;”生成若干个统一形式的 GRIB2 文件,并传输至在线磁盘存储,供数据服务门户系统检

索调用。

3.3 归档存储

针对数据在线存储方式,参考设计的统一文件组织形式“s2s_[centre]_ \${cyear} _[dataDate] _[shortName]. grib”,建立目录 \${DATADIR}/[centre]/[cyear]/[year]/[month]/[day],将对应的数据存储于该目录中。其中,DATEDIR 为预定义的目录,year,month,day 分别是 dataDate 属性中抽取出的对应字段。这种文件组织形式和存储目录设计,可直观的由文件名推知文件存储路径,利于数据高效检索获取,也可支持 OPeNDAP 协议的数据访问。

针对 MARS 归档入库处理,可直接对收集到的数据解析生成一个入库请求文件,进行入库操作。如果收集的数据不是以单层次类型组织,则需经过分离处理后,再逐一入库。分离处理基于 grib_filter 命令实现,处理流程如图 2。

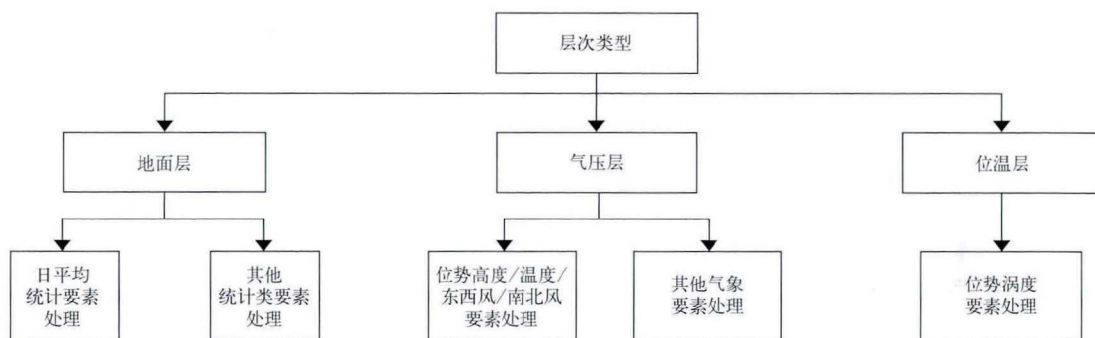


图 2 GRIB 数据分离处理流程

Fig. 2 The flow of splitting GRIB data

S2S 回算集合预报数据一般涉及超过 10 年的数据,而且有固定静态配置预报和动态配置预报之分,数据量巨大。为了高效检索所需数据,针对回算集合预报数据,在应用 MARS 存储管理数据时,未采用通常形式的 MARS 树节点设计(树节点按年月日设计,如 2015:201512:20151201,超立方体结构的叶子节点不包含日期属性),而是针对性的优化设计,以减少索引数量。对于固定静态配置预报数据,采用删除日期的年月形式 MARS 树节点设计,如 2010:201001,叶子节点中包含日期属性,其取值为指定月份具有数据的所有日期,如 20100110,20100120,20100131;对于动态配置预报数据,

MARS 树节点采用数据生产日期形式的 MARS 树节点设计,如 20150101,而叶子节点则包含这日生成的所有回算数据日期,如 19950101,19960101,19970101,19980101,……,20140101。

3.4 数据服务门户

中国气象局 S2S 数据服务门户网站作为 S2S 数据归档中心的一部分,主要针对国际用户,故未提供中文页面。此门户网站提供对各中心模式及气象要素的详细描述、简单帮助信息、数据检索下载和后台数据统计功能。通过设计气象要素表、气象要素属性表、模式表、模式属性表、各中心提供的气象要素表,网站上数据下载、模式及要素说明的页面配置

动态展示予以实现。针对用户的数据服务主要是统一实时和回算预报数据提供一致的检索下载服务,按中心、变量目录方式及自由文本搜索方式下载,支持原始 GRIB2 格式和经在线转换的 NetCDF 格式数据获取,并采用订单方式的下载处理,亦可批处理下载多个文件数据。数据下载页面中,有效数据时

间范围通过配置或查询动态获取。自由文本搜索方式检索基于 Apache Lucene 全文搜索实现。中国气象局 S2S 数据服务门户网站下载页面见图 3。

用户下载的 S2S 数据主要应用于多模式预测及验证、极端天气事件(如 MJO 事件)、季风模拟、遥相关等方面的研究。



图 3 中国气象局 S2S 数据服务门户网站下载页面
Fig. 3 Download page of CMA S2S data portal

4 小 结

本文介绍了各 S2S 生产中心提供的数据,并在此基础上分析数据特征,结合归档中心建设需求,重点针对数据交换、同步、存储管理难点问题,从系统设计和关键技术实现方面进行讨论。中国气象局 S2S 数据归档中心特征如下:

1) 归档中心全流程建设涉及到数据交换、数据质量检查及处理、归档存储和数据服务 4 个流程的处理,实现数据的完整收集、归档及对外服务。

2) 数据交换主要基于 FTP 传输数据,记录日志获取缺失数据列表,通知补传并结合 WebAPI 方式主动补调缺失数据实现同步。

3) 数据质量控制及处理基于 GRIB2 编码特征,主要应用 GRIB_API 工具命令实现。

4) 归档中心管理的数据包括近实时集合预报数据和回算预报数据,全部在线存储,结合特定的目录及文件名设计,实现高效的数据检索下载。

目前,中国气象局 S2S 归档中心处于运行状态中,未来会考虑进一步完善数据质量检查功能,部署更完善的数据同步策略并及时更新 S2S 数据服务门户网站,针对国内用户提供更多的数据服务方式。

随着集合预报向着更高分辨率、更多集合成员数、更高预报频次的方向发展,更大数据量的集合产品将会产生,要求具有更强性能的存储系统,并能适应用户需求灵活快捷的提供经过定制、裁剪的数据,减少不必要的数据传输。国际上此类集合预报数据管理和服务趋势主要有两大类:一类是基于 OpenDAP (Open-source Project for a Network Data Access Protocol) 协议,结合分布式服务和搜索技术,使用全部在线或在线及离线存储,应用简单程序工具通过命令或 Web 方式提供定制化的服务;另一类是以 ECMWF 的 MARS 为代表,针对特定数据类型采用专有数据管理技术并结合分布式技术,自动化处理在线、离线数据,通过命令、Web 方式及 WebAPI 方式提供定制化服务。近年来国内随着大数据云化技术逐步应用,一方面将大数据部署在云系统上,另一方面对上层提供大数据服务的云化,亦对气象数据的管理服务提供了机会。当前 CIMISS (China Integrated Meteorological Information Service System, 全国综合气象信息共享系统) 已业务化,并考虑应用云计算、大数据、物联网、移动互联等技术,依托专有云和公有云的混合云架构技

术进行升级。S2S 数据作为一类国际项目合作并主要面向科研应用的气象数据,S2S 数据归档中心建设亦可较好地融入 CIMISS 系统,实行统一的管理和服务,满足各类用户的需求。

参 考 文 献

- [1] Hoskins B. The potential for skill across the range of the seamless weather-climate prediction problem: A stimulus for our science. *Quart J Roy Meteor Soc*, 2013, 139: 573-584.
- [2] Vitart F, Ardilouze C, Bonet A, et al. The sub-seasonal to seasonal prediction (S2S) project database. *Bull Amer Meteor Soc*, 2017, 98: 163-173, DOI: 10.1175/BAMS-D-16-0017.1.
- [3] 齐艳军, 容新尧. 次季节-季节预测的应用前景与展望. *气象科技进展*, 2014, 4(3): 74-75.
- [4] 任宏利, 吴捷, 赵崇博, 等. MJO 预报研究进展. *应用气象学报*, 2015, 26(6): 658-668.
- [5] Bougeault P, Toth Z, Bishop C, et al. The THORPEX interactive grand global ensemble (TIGGE). *Bull Amer Met Soc*, 2010, 91: 1059-1072.
- [6] Richard S, Masayuki K, Piers B, et al. The TIGGE project and its achievements. *Bull Amer Meteor Soc*, 2016, 1, DOI: 10.1175/BAMS-D-13-00191.1.
- [7] 卞晓丰. 交互式全球集合预报产品数据管理应用研究. *科研信息化技术与应用*, 2013, 4(6): 64-70.
- [8] 沈文海, 赵芳, 高华云, 等. 国家级气象资料存储检索系统的建立. *应用气象学报*, 2004, 15(6): 727-736.
- [9] 李集明, 沈文海, 王国复. 气象信息共享平台及其关键技术研究. *应用气象学报*, 2006, 17(5): 621-628.
- [10] 高峰, 王国复, 喻雯, 等. 气象数据文件快速下载服务系统的设计与实现. *应用气象学报*, 2010, 21(2): 243-249.
- [11] 马渝勇, 徐晓莉, 宋智, 等. 省级气象信息共享系统的设计与实现. *应用气象学报*, 2011, 22(4): 505-512.
- [12] 李永生, 曾沁, 徐美红, 等. 基于 Hadoop 的数值预报产品服务平台设计与实现. *应用气象学报*, 2015, 26(1): 122-128.
- [13] 熊安元, 赵芳, 王颖, 等. 全国综合气象信息共享系统的设计与实现. *应用气象学报*, 2015, 26(4): 500-512.
- [14] WMO. Guide to the WMO Table Driven Code Form Used for the Representation and Exchange of Regularly Spaced Data In Binary Form; FM 92 GRIB Edition 2. [2016-07-15]. http://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB_062006.pdf.
- [15] Access MARS-ECMWF WebAPI. [2016-07-15]. <https://software.ecmwf.int/wiki/display/WEBAPI/Access+MARS>.
- [16] 肖华东, 孙婧, 张玺, 等. MARS 软件在数值预报模式产品数据管理中的应用. *应用气象学报*, 2016, 26(2): 247-256.
- [17] ECMWF Operations Department. MARS Use Guide. [2016-07-15]. <http://www.ecmwf.int/publication/manuals/mars/guide/userguide.pdf>.
- [18] ECMWF S2S Home. [2016-07-15]. <https://software.ecmwf.int/wiki/display/S2S/Project>.

Design of CMA S2S Data Archive Center and Key Technology

Xiao Huadong¹⁾²⁾³⁾ Sun Jing¹⁾ Sun Chaoyang⁴⁾ Nie Yuanding¹⁾ Zhao Chunyan¹⁾

Guo Feng¹⁾ Chang Biao¹⁾ Zhang Xinnuo¹⁾ Liu Liming¹⁾

¹⁾ (National Meteorological Information Center, Beijing 100081)

²⁾ (Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077)

³⁾ (University of Chinese Academy of Sciences, Beijing 100049)

⁴⁾ (National Climate Center, Beijing 100081)

Abstract

S2S, sub-seasonal to seasonal prediction project, is a WWRP/THORPEX-WCRP joint research project established to improve forecast skill and understanding on the sub-seasonal to seasonal time scale, and promote its uptake by operational centers and exploitation by the applications community. To achieve many of these goals, the establishment of an extensive data base of sub-seasonal (up to 60 days) forecasts and reforecasts (sometimes known as hindcasts) is advocated. Managing an extensive data base for the sub-seasonal time scale is a challenging task for the diverse characteristic of these forecasts (e. g. , start dates, length of the forecasts, average periods, and updating frequency of forecasts).

Open access to forecast data and user interface friendly data bases are important requirements for construction of S2S archiving data. CMA, one of two S2S data archive centers, is responsible for collecting S2S data from all data providers, performing basic quality check, archiving and storing data, and providing data service. The traditional FTP and a simple data downloading program based on ECMWF WebAPI is used to collecting data provided by all eleven S2S data centers. The data synchronization with another archive center of ECMWF is implemented by applying the combination strategies of common routine data transfer through FTP, ECMWF WebAPI based data download for missing reforecast data, and transferring missing real-time data copied from tapes directly by MARS system command. The file of compressed format TAR and GRIB2 is checked by executing command of tar and grib_get and its return value. A program based on the tigge_check command provided by GRIB_API is also developed to check the attribute of GRIB2 data, since all S2S data are saved as GRIB2 format. Through using GRIB_API command of grib_filter and designing a unified data form organized by a single center, a single date and a single meteorological parameter field, the unified S2S data are created to store in the online storage system which makes fast retrieval of data possible and open access. In the meantime, the unified form of S2S data file could be easily obtained their real storage addresses by their file names. The S2S data portal provides both free text and faceted search method to access forecast and reforecast data in the format of GRIB2 and NetCDF through online converting the data format.

CMA S2S archive center is in operation since 15 November 2015. With the availability of data of all data providers, the construction of S2S archive center will be continued to improve the in-time of data integrity and data access methods.

Key words: S2S; ensemble forecast; hindcast; on the fly