

# 奥运短时临近预报实时数据处理

王玉彬<sup>1</sup> 周海光<sup>2</sup> 余东昌<sup>1</sup> 周 勇<sup>3</sup> 苏德斌<sup>1</sup> 梁 丰<sup>1</sup>

(1. 北京市气象局, 100089; 2. 中国气象科学研究院灾害天气国家重点实验室;  
3. 国家气象信息中心)

**提 要:** 详细分析了目前短时临近预报系统对自动气象站数据质量控制、探测数据更新频率、组网雷达时间同步观测与基数据实时传输、探测数据的实时传输技术、基于目录监控的数据监控与分发机制、基于开源的软件开发、数据格式等方面的需求。在此基础上,建立了北京奥运短时临近预报数据快速收集处理系统并已经投入业务运行,该系统实现了探测数据的实时传输、组网雷达时间同步观测和基数据实时传输、基于消息驱动的数据实时监控和分发、探测数据格式规范化等功能,基本满足北京2008年奥运短时临近预报系统的数据预处理业务需求。

**关键词:** 数据传输 实时处理 消息驱动 自动气象站 质量控制 临近预报 雷达基数据

## Real Time Data Processing Technique on Very Short-Range and Nowcasting for Beijing 2008 Olympic Games

Wang Yubin<sup>1</sup> Zhou Haiguang<sup>2</sup> Yu Dongchang<sup>1</sup> Zhou Yong<sup>3</sup> Su Debin<sup>1</sup> Liang Feng<sup>1</sup>

(1. Beijing Meteorological Bureau, 100089; 2. State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences; 3. National Meteorological Information Center)

**Abstract:** The data processing kernel problems of the very short-range and nowcasting are analyzed in detail. These include quality control of automatic weather station data, data update frequency, network Doppler radar synchronization observation and level-II data transfer in time, detecting data transfer in real time, detecting data monitoring and dissemination based on the directory monitoring and message drive, software development based on the open source technology, and data format etc. Some techniques are adopted to meet the short-term and nowcasting demands. Based on these techniques, the DRAPS (Data Rapid Acquisi

资助项目: 科技部国家科技攻关计划“北京奥运短时临近预报实时业务系统建设”(编号: 2005BA904B05), 北京市科委科技计划“北京奥运会国际天气预报示范计划支持技术研究”(编号: Z006279040191)

收稿日期: 2008年1月28日; 修定稿日期: 2008年2月2日

tion and Process System) of nowcasting for Beijing 2008 Olympic games was developed and run in operation more than one year. Some advanced techniques were adopted in DRAPS, such as data transfer in real time, network radar synchronization observation, data monitoring and dissemination in time based on message driven, and data format standardization etc. The system basically meets the demands on the data processing for very short-term and nowcasting.

**Key Words:** data transfer real-timely processing message driving AMS data quality control nowcasting radar data

## 引 言

大气探测数据的快速收集和处理是气象业务的基础,是短时临近预报的关键环节之一,国内外已经进行了大量的研究,目前已经有很多研究成果投入业务应用,极大地促进了气象业务现代化<sup>[1]</sup>。美国从 1980 年代就开始本地数据收集和分发(LDAD: Local Data Acquisition and Dissemination System)的研究与开发,目前 LDAD 是美国国家天气局(NWS)收集、处理本地数据和分发气象数据的基本支撑业务平台<sup>[2-3]</sup>。LDAD 系统包括了探测数据监视、数据解码、质量控制、数据分发等功能。为了数据的安全,LDAD 将探测数据环境分为内网和外网,这样就增加了数据处理的时间,不利于临近预报系统对实时数据的需求。

近些年,随着各种气象探测设备的大规模布网以及气象探测数据在短时临近预报中的不断深入应用,短时临近预报业务对本地探测数据的实时收集、自动气象站数据实时质量控制、多种探测资料的时间同步、数据更新频率、数据格式和数据组织方式等提出了新的要求。虽然目前的气象传输网络速度较快,但由于探测数据种类繁多,而且有些数据量较大,通信网络远远不能满足大容量数据的实时传输<sup>[4-6]</sup>。各种探测数据的时间分辨率也不一样,例如自动气象站数据时间分辨率最高为 1 分钟,风廓线雷达数据时间分辨率为 6 分钟,新一代多普勒天气雷达 VCP21

体扫模式时间分辨率为 6 分钟左右。如何将这些时间分辨率不一致的探测数据进行时间同步处理,是数据处理系统需要解决的关键问题之一。各种探测数据的格式不统一或不标准化也是阻碍探测数据的应用和短时临近预报系统研发的瓶颈问题。

上述问题国内开展的研究很少,如何解决这些问题,并在此基础上建立符合短时临近预报需求的探测数据实时采集、处理系统,是提高探测数据利用水平、提高短时临近预报准确率的重要环节。本文拟在这些方面开展分析研究,建立满足短时临近预报业务需求的探测数据实时加工、处理系统,也为今后全国范围的相关研究提供参考。

## 1 自动气象站数据质量控制

目前自动气象站的数据质量参差不齐,如何对自动气象站数据进行质量控制工作,将其应用到短时临近预报业务中,是一件非常紧迫的任务。

气象学家很早就对自动雨量站的数据完整性、可靠性和准确性等进行了详细的理论研究<sup>[7-15]</sup>。王颖等<sup>[13]</sup>根据 2000 年自动站和人工观测气温资料,对两种气温序列进行了对比分析。研究表明自动站观测气温与历史序列无显著性差异;自动站观测数据可以用于业务,但需进行一定时间的平行观测。张强等<sup>[14]</sup>利用 19 种不同型号的自动雨量站实测对比观测资料,进行降雨资料质量评估,提出了采用基于各自动雨量站观测结果的拟合

降雨量作为降雨量参考标准的新思路,并对拟合降雨量的可行性、适用性以及计算方法进行了详细论述。但如何对自动气象站资料进行实时质量控制,将其用于短时临近预报,这方面的研究还很少。作者在这方面进行了一些探索和研究,并在此基础上建立自动站数据实时质量控制软件系统。一年多的业务试运行表明,该软件系统能够满足短时临近预报系统对自动站数据质量控制的实时要求。鉴于临近预报对自动站质量控制算法的实时性要求很高,主要参考世界气象组织关于自动站质量控制的指导性文件<sup>[12]</sup>,并根据探测数据的具体特点以及算法实时性的要求,对其进行了改进。

自动气象站的数据质量控制主要包括下述 4 个方面。(1)原始要素范围检查:自动站探测数据必须在一个给定的合理范围之内变动;(2)时间一致性检查:对于一个给定探测量,其相邻两次的探测值必须小于一个给定的数值;(3)瞬时值变化的最小值检查:对于一个给定探测量,其相邻 1 小时的测量值必须大于某一个给定值;(4)内部一致性检查,对于同

一个时刻不同气象要素值应该满足一定的关系。上面的算法比较全面地考虑了自动气象站数据质量控制的各个环节,同时从各种测量数据的物理关系角度对其进行了综合处理,能够比较真实地再现数据的实际状态。

## 2 数据传输技术

### 2.1 组网雷达时间同步与雷达基数据实时传输

由于目前多普勒天气雷达均使用 RDA 计算机的时钟作为雷达系统的授时系统,因此多部雷达在组网观测时,每部雷达系统使用各自的 RDA 计算机时钟作为参考时间,很难保证多部雷达系统的时间同步。基于苏德斌、张建云针对提高多部雷达基数据的时间一致性提出的区域雷达网雷达实时同步扫描方案,对于多普勒天气雷达数据传输,本文采用 GPS 授时、雷达扫描分片传输技术,解决了多部雷达大数量实时雷达数据传输与处理问题。图1是组网多普勒雷达时间同步观测

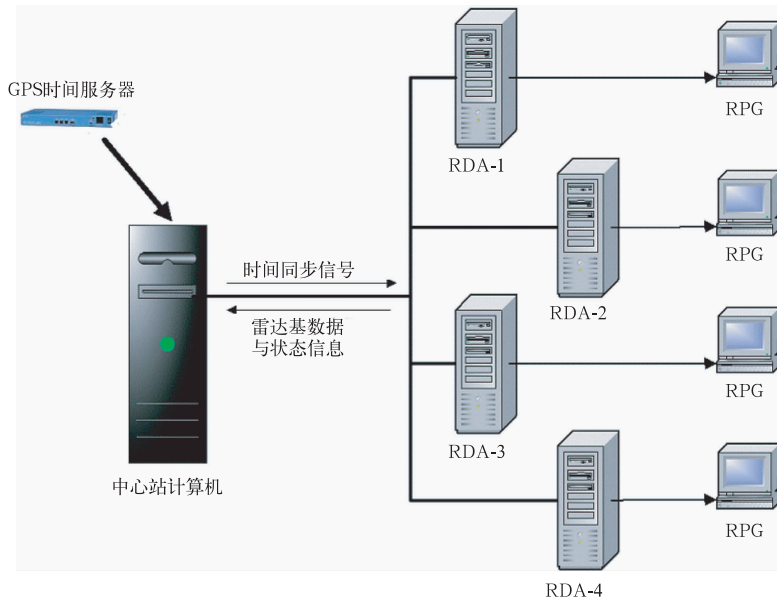


图 1 组网多普勒雷达时间同步观测与基数据实时传输原理图

RDA: 雷达数据采集子系统;RPG: 雷达产品生成子系统

与基数据实时传输原理图。GPS 授时多部雷达时间同步技术是指:在中心站建立 GPS 时间授时系统,在每部雷达 RDA 计算机上建立 GPS 时间同步进程,GPS 授时系统每隔一定时间间隔向每部雷达 RDA 计算机广播时间信息,每部雷达 RDA 计算机据此同步各自的时钟,确保多部雷达时间同步;因此多普勒雷达基数据和 RPG 生成的二次产品中的时间信息也同 GPS 授时系统广播的时间相一致。1990 年代,美国在多基地天气雷达外场试验中也是使用 GPS 授时技术确保主动雷达和被动接收天线系统之间的时间同步<sup>[16]</sup>。

多普勒天气雷达资料是临近预报中使用频次最高的资料之一,由于临近预报对资料的实时性要求很高,而雷达基数据的数据量较大且时间分辨率高,这对数据传输提出了很高的要求。以 CINRAD/SA 多普勒天气雷达为例,雷达完成一个 VCP21 模式的体积扫描需要 6 分钟左右,基数据体扫文件为 9.4M 左右,而 CINRAD/CC 和 CINRAD/CD 多普勒天气雷达的体扫文件由于基数据格式的差异,其数据量更大。采用数据压缩技术可以大幅度地减少其数据尺寸,但压缩技术很耗时,影响处理时效;而且压缩效率同雷达探测时的天气状况密切相关,特别是当有大范围的降水过程时,压缩后的数据尺寸依然比较大。

目前我国气象业务网中雷达基数据的传输方式是当雷达完成一个体扫之后才开始将基数据通过网络传输,此时对于 VCP 21 模式的雷达基数据而言,传输时间已经延迟了 6 分钟左右,显然不能满足临近预报数据实时传输要求。虽然目前网络带宽较高,但若一个中心站同时接收多个雷达站的体扫文件,还会造成网络的短时拥挤;在没有雷达数据传输时,又造成大量网络带宽的闲置和浪费。

鉴于此,作者采用多普勒天气雷达基数据实时同步传输技术,即雷达完成一根径向

数据处理后,就将此根径向数据传输到中心站,由中心站将径向数据直接写入中心站雷达基数据文件。当雷达开始体扫时,在中心站便直接生成一个该雷达的基数据文件(最初为空文件),随着体扫的进行,该雷达的各径向数据由雷达站顺序逐一陆续传到中心站,并直接写入该已建好的基数据文件中,当雷达完成一个体扫之后,雷达基数据也传输完毕。该技术在北京及周边地区 4 部雷达同步组网传输系统中,每部雷达的一个 VCP21 体扫资料传输仅需要 220Kbps 的带宽,极大地降低了对网络带宽的需求,同时实现了雷达基数据真正意义上的实时同步传输。由于临近预报中雷达数据的处理计算量很大,该技术也为临近预报处理预留了提前量。北京敏视达雷达有限公司已经通过软件编程在 CINRAD/SA 雷达上实现了上述功能。

## 2.2 探测数据传输流程

根据探测设备地理位置及短时临近预报业务系统对探测数据空间范围的需求,将探测数据划分为本地及省际两个区域。对于本地探测数据首先传输到本地气象信息中心,再由信息中心分发给各个业务系统。不同的探测数据可以采取不同的传输解决方案,表 1 给出了各种资料的传输方式。

表 1 各类资料的传输方式

序号	资料类型	传输方式
1	雷达数据资料	专线
2	风廓线雷达资料	专线
3	自动气象站资料(1,2 级站)	专线
4	自动气象站资料(其他类型)	CDMA/GPRS
5	闪电定位资料	GPRS
6	其他资料	局域网

省际数据传统业务传输模式是基于国家气象信息中心的实时库,采用文件接口或数据库接口进行开发。文件接口具有业务化水平高、稳定可靠、易实现等优点。传统业务模

式下,数据传输流程多采用集中传输模式,也即将所有站的数据汇总后再传输。数据从测站到达临近预报业务系统需要经过测站逐级上传至国家气象信息中心,然后再分发至临近预报系统所属区域的信息中心。这些中间环节的处理大部分采用定时启动方式,非常容易造成网络延时,不能满足时效要求;而且这种传输模式也容易造成网络瞬间传输高峰和数据服务器的处理高峰。

可以看出,传统传输模式是:测站资料首先逐级上传,而不是共享使用;即便测站邻近的用户,也需要待资料上传至国家级后再行下发,才可获得邻近测站的数据。而新方法则是在测站观测资料逐级上传的同时,便进行共享使用。为了提高探测数据的传输时效,可以将省际自动气象站资料和雷达资料等数据直接由观测站点传输到本地数据共享平台并立即分发给业务系统。作者已经将新的数据传输流程成功地业务化,新流程基于北京区域中心宽带网和北京区域数据共享平台,将自动气象站资料、北京及周边地区4部多普勒雷达资料由观测站点直接传输到区域数据共享平台后立即分发给各业务系统使用。与传统模式相比,新流程减少了大量中间环节,极大地提高了传输时效。以自动气象站数据为例,采用传统方式,数据从观测到最终可用至少需要大约312秒,超过自动气象站观测的时间间隔;而在2007年8月1—31日北京奥运气象服务演练数据传输测试中,新的流程可以在180秒以内收集到90%以上站点的资料,5分钟以内到达的数据达97.65%。北京及周边地区4部雷达数据的正点到达率为99.08%,缺失的数据主要是由雷达故障和系统维护所导致。

### 2.3 基于目录监控和消息驱动的数据实时监控、收集、处理与分发

探测数据具有时间分辨率高、数据更新

速度快的特点,数据收集与处理系统必须对探测数据进行实时监控,当有新的探测数据到达时,就应启动数据处理模块对新到达的数据处理与分发。以前的数据监控都是采用文件与目录定时扫描技术,即监控模块定时扫描用户指定的一组数据目录,当有新的数据到达时就启动后续处理进程。由于探测数据更新速度非常快,监控模块需要非常频繁地扫描磁盘,这种技术体制浪费了宝贵的计算资源,使得服务器负载过重,处理时效性较差。

为了提高计算资源的利用率和响应速度,作者研发了基于消息驱动的数据实时监控、收集、处理与分发子系统,明显地降低了服务器负载,增强了系统的实时处理能力。消息驱动是指软件系统各个模块间的组织和调用方式,这种方式可以提高运行模块对相应事件的响应速度,加快处理流程。该子系统是在Linux系统上,利用Inotify机制实现的。Inotify是在Linux 2.6.13内核版本中引入的新功能,采用事件驱动机制,为用户监视文件系统的变化提供了强大的通知支持功能。它的主要特点是:不仅可以监视文件,还可以监视目录;使用系统调用通知文件系统的变化事件;由于使用文件描述符作为接口,因而可以使用通常的文件I/O操作来监视文件系统的变化;如果被监视目标在可移动介质上,在umount该介质上的文件系统后,被监视目标对应的watch将被自动删除,并且会产生一个umount事件。目前可以使用C语言调用Inotify函数库实现其所有功能,但编程比较复杂。Pyinotify是建立在Python 2.3基础之上的基于Inotify的文件系统监视模块<sup>[17-18]</sup>,利用Pyinotify对文件系统进行监视时,编程简单、容易实现、功能强大。

作者综合使用Pyinotify技术和消息驱动技术实现数据的实时监控、收集、处理与分发。利用Pyinotify技术对一组用户定义的

探测数据目录在后台进行监视,当有新的探测数据到达时,Pyinotify 自动生成文件系统变化通知消息,系统调用相应的数据处理进程对探测数据进行处理和分发。系统不需要定期扫描磁盘,有效地降低了服务器负载,提高了数据处理效率。

### 3 DRAPS 系统设计与实现

基于上述分析,在科技部国家科技攻关计划“北京奥运短时临近预报实时业务系统建设”项目的资助下建立了北京奥运短时临近预报数据快速收集处理系统(DRAPs: Data Rapid Acquisition and Process System),该系统是 2008 年北京奥运会临近预报业务系统的重要组成部分。DRAPS 是一个基于目录监控和消息驱动机制实现的数据快速收集与处理系统,系统综合采用了 XML、多线程等技术,并且全部采用开放源代码平台,采用 Python 开发。系统包括北京及其周边地区探测数据的实时采集、传输、处理、质量控制与分发等功能,为北京奥运短时临近预报业务系统实时提供经质量控制的自动气象站数据、风廓线仪数据、闪电定位数据、多普勒天气雷达基数据及产品数据、卫星产品数据、数值预报产品等。该系统满足世界气象组织北京 2008 年奥运会预报示范计划(B08FD)中短时临近预报业务系统以及北京市气象局临近预报业务平台对数据实时处理的业务需求。

#### 3.1 探测数据更新频率

探测数据更新频率的确定需要对短时临近预报业务需求、探测系统的探测能力、通讯网络传输能力、计算机系统处理能力等因素进行综合考虑,目前系统确定的资料数量和更新频率见表 2。

表 2 探测资料更新频率

序号	资料类型	数量	更新频率
1	自动气象站/雨量站	106 个	5min
2	风廓线雷达	3 部	6min
3	闪电定位监测	1 套	实时
4	多普勒天气雷达	4 部	6min
5	静止卫星观测	2 颗	30min
6	GPS/MET 水汽资料	/	30min
7	数值预报产品	/	3h

#### 3.2 探测数据格式

目前各种气象探测设备的数据格式种类繁多,每个短时临近预报业务系统对输入数据格式又有着特定的要求,这使得探测数据的读取非常繁琐。目前探测数据格式有两种解决方案:即采用探测设备自身定义的格式,并提供解码说明或解码程序;或提供国际通用的数据格式。对于自动气象站数据、闪电定位数据、数值预报产品数据,本系统均采用 NetCDF 通用数据格式,为短时临近预报业务系统提供符合国际惯例的数据环境。临近预报系统可以直接使用本系统分发的各种探测数据,其探测数据读取模块不需做修改,使本系统很容易同其他系统进行无缝集成,为系统的推广应用奠定了很好的技术基础,同时也加快了临近预报系统的本地化部署速度。在 2006 年和 2007 年世界气象组织北京 2008 年奥运会预报示范计划中,来自中国、澳大利亚、加拿大、中国香港、美国等 8 个短时临近预报业务系统就是直接使用本系统分发的各类探测数据进行临近预报,北京市气象局临近预报业务平台也使用本系统作为数据输入平台。

#### 3.3 基于开源架构的软件开发

采用开放源码软件开发架构实现数据的收集、监控、处理与分发,采用的主要开源软件有 python 编程语言,目录监控部分运行于 Linux 操作系统。采用开源架构可以降低

系统的开发成本和运行成本,提高系统的可维护性、通用性和可扩展性。为了有效地降低服务器负载,提高系统处理时效,采用基于 Pyinotify 机制的目录监控和文件系统变化消息驱动方式实现探测数据实时收集、处理、存储与分发。

### 3.4 HTKDRAPS 系统的体系结构

图 2 是 DRAPS 系统的体系结构图,系统由多台数据收集前置机和数据处理机组成,数据收集前置机主要负责北京市气象局所属观测站网和北京周边地区(省级)观测站

网的探测数据收集,目录监控进程运行在数据收集前置机上实现数据文件的监控与分发进程调用等功能。数据处理机完成数据的质量控制、格式转换、要素分解和分发等功能。消息传送路径如图 2 所示,F1 表示数据采集端目录监控进程接收到新文件到达消息后立即启动文件发送进程,将文件发送至数据处理服务器;F2 表示目录监控进程通知数据处理服务器上的消息处理进程有新文件到达;F3~F6 则表示消息处理进程接收 F2 消息后,根据预先定义启动格式转换、分发、质量控制或要素分解等进程。

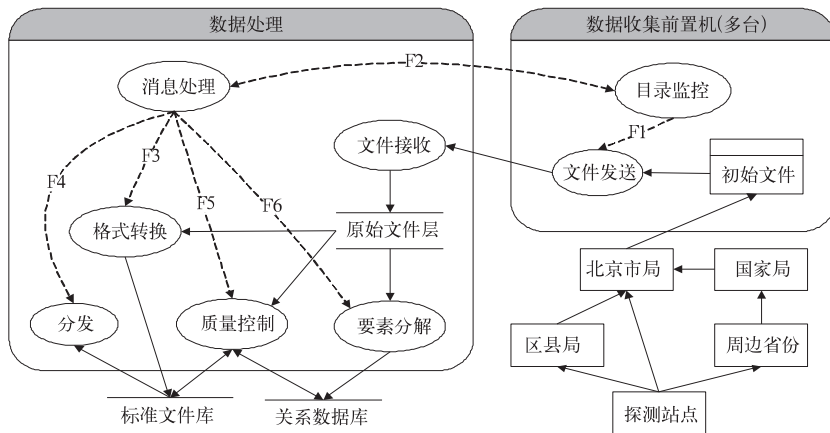


图 2 DRAPS 系统的系统体系结构图

实线为探测数据的可能传输路径,虚线为消息传递路径

## 4 结 论

作者对自动气象站数据的实时质量控制、探测数据的实时收集、组网雷达时间同步、临近预报数据更新频率、临近预报数据格式、数据文件组织方式等进行了分析研究。在此基础上建立了探测数据快速收集处理系统,该系统能够满足短时临近预报系统对探测数据的实时性要求,有效地提高了网络资源利用率、探测数据的利用率;优化了探测数据的组织管理方式,使之更符合短时临近预

报业务系统的需求。

(1) 在国内天气雷达组网观测中,首次应用 GPS 授时技术确保多部雷达时间同步;在国内业务系统中首次引入适合多普勒天气雷达基数据实时传输的径向数据传输方案,实现了高时间分辨率的雷达基数据的实时传输,能够满足短时临近预报系统对雷达数据的实时需求。

(2) 在业务系统中首次引入基于 Pyinotify 技术的消息驱动方式的探测数据实时收集、处理、处理与分发,实现了网络和计算服务器的负载均衡,提高了系统对数据实时处

理的效率。

(3) 采用开放源码软件技术实现数据的收集与处理,降低系统的运行成本,也有利于系统的维护和升级。

(4) 采用国际通用数据格式规范统一了本地探测数据格式,提供了符合国际惯例的数据环境,使本系统很容易同其他系统进行无缝集成,为系统的推广应用奠定了很好的技术基础。

**致谢:** 本论文的选题和设计是在北京市气象局王建捷研究员的指导下进行的,北京市城市气象研究所王迎春、谭晓光研究员对本项研究课题提出了宝贵建议。敏视达雷达有限公司总经理张建云在雷达数据实时处理技术方面给予了悉心指导。同时,要感谢本课题组其他成员:窦以文、季崇萍、刘国红、刘珂,他们为这篇论文提供了很多帮助。

## 参考文献

- [1] 王玉彬,周海光,苏德斌,等. 天气预警系统技术基础及设计[M]. 北京:气象出版社,2006.
- [2] <http://www-sdd.fsl.noaa.gov/~ldad/>
- [3] <http://www.fsl.noaa.gov>
- [4] 潘新民,吴富山. 多普勒天气雷达远程数据传输同步处理系统[J]. 气象,1999,25(3):34-37.
- [5] 吕刚,华连生,钱毅. GSM 网络通讯技术在雨量遥测中的应用[J]. 气象,2005,31(5):84-86.
- [6] 陈少平,田刚,王建国. 三峡自动气象站实时资料数据库及其应用[J]. 气象,2005,31(12):69-72.
- [7] Rinehart R E. Out of level instruments: errors in hydrometeor spectra and precipitation measurements [J]. J Climate Appl Meteor,1983,22:1404-1415.
- [8] Sevruk B, Hamon W R. International Comparison of national precipitation gauges with a reference pit gauge[R]. 1984; WMO/TD No. 38.
- [9] WMO. Guide on the global data processing system [R]. 1993; WMO-No. 305,VI. 1-VI. 27.
- [10] Goodison B E. WMO solid precipitation measurement inter-comparison in papers presented at the WMO technical conference on instruments and methods of observation (TECO294)[C]. WMO/TD No. 588. Geneva:1994:15-19.
- [11] WMO. Guide to meteorological instruments and methods of observation[R]. 1996; WMO-No. 8,1-12.
- [12] Igor Zahumensky. Guidelines on quality control procedures for data from automatic weather stations. World meteorological organizations commission for instrument and methods of observation OPAG-surface expert team on surface technology and measurement techniques[C]. Switzerland, 13-16 October 2004, CIMO/OPAG-SURFACE/ET ST&MT-1/ Doc. 6. 1(2).
- [13] 王颖,刘小宁. 自动站与人工观测气温的对比分析[J]. 应用气象学报,2002,13(6):741-748.
- [14] 张强,涂满红,马舒庆,等. 自动雨量站降雨资料质量评估方法研究[J]. 应用气象学报,2007,18(3):365-372.
- [15] 任芝花,熊安元. 地面自动站观测资料三级质量控制业务系统的研制[J]. 气象,2007,33(1):19-24.
- [16] 周海光. 多普勒天气雷达风场反演技术研究进展[J]. 地球科学进展,2002,17(3):384-390.
- [17] <http://pyinotify.sourceforge.net/>
- [18] <http://www.kernel.org>