

# L波段探空秒级资料稀疏化方案及检验方法

李庆雷<sup>1</sup> 远芳<sup>1</sup> 杨贵<sup>2</sup> 廖捷<sup>1</sup> 胡开喜<sup>1</sup> 姚爽<sup>1</sup> 周自江<sup>1</sup>

(1 国家气象信息中心, 北京 100081; 2 内蒙古自治区二连浩特市气象局, 二连浩特 011100)

**摘要:** 高垂直分辨率探空资料的稀疏化是资料同化应用过程中非常重要的技术环节。通过甄别L波段秒级探空不同气象要素的时空变化拐点, 提出一套自适应的稀疏化方案。该方案可以显著增加温度、湿度、风特性层数目, 使得单次探空廓线的层数较现行台站业务软件提供的多近200%, 既能精准地描述气象要素垂直变化特征, 又能有效地被再分析同化系统应用。应用基于模式背景场(ERA-Interim再分析资料)的观测资料偏差评估技术, 通过对比分析不同来源的探空要素数据量、偏差(Bias)和均方根误差(RMSE)等指标, 对稀疏化方案的效果进行了量化检验。结果表明: 该稀疏化方案不仅实现了再分析同化应用过程中的探空廓线层数的显著增加, 而且其Bias和RMSE的大小及垂直分布较已有整合廓线数据更合理。该稀疏化方案为更有效地将我国高垂直分辨率L波段探空资料应用于大气再分析提供了必要保证。

**关键词:** 秒级探空, 稀疏化, 量化检验, 再分析, 同化

DOI: 10.3969/j.issn.2095-1973.2018.01.017

## A Sparsification Scheme and Evaluation of the L-Band Radiosonde High-Resolution Data

Li Qinglei<sup>1</sup>, Yuan Fang<sup>1</sup>, Yang Gui<sup>2</sup>, Liao Jie<sup>1</sup>, Hu Kaixi<sup>1</sup>, Yao Shuang<sup>1</sup>, Zhou Zijiang<sup>1</sup>

(1 National Meteorological Information Centre, Beijing 100081

2 Erenhot Meteorological Bureau of the Inner Mongolia Autonomous Region, Erenhot 011100)

**Abstract:** This paper puts forward a suitable sparsification scheme for L-band second-level radiosonde data. The sounding vertical profile may describe more accurately the vertical change of meteorological elements with increased number of the significant layers. The background field (ERA-Interim reanalysis) is applied for bias evaluation of the observation data. Sounding data from different sources are utilized for comparison and analysis. The bias, the root mean square error (RMSE) and other quantitative indicators are calculated to assess the effect of the sparsification scheme. The results show that this scheme can not only increase the numbers of the radiosonde profile levels, but also achieve better Bias and RMSE results. This scheme may provide necessary guarantee for the assimilation applications of the high-resolution L-band radiosonde data in China for the future.

**Keywords:** second-level radiosonde, sparsification, quantitative evaluation, reanalysis, assimilation

### 0 引言

常规探空观测资料因其质量稳定可靠、垂直层数多、能够准确描述大气三维结构, 而成为提供陆地大气状态极其重要的信息来源。这使得探空资料在数值预报<sup>[1]</sup>、天气分析<sup>[2-3]</sup>、气候变化<sup>[4]</sup>、卫星资料校准<sup>[5]</sup>等研究中发挥着重要作用<sup>[6]</sup>, 尤其是在数值预报系统进行资料同化过程中, 探空资料更是成为改进模式初始场质量、提高预报精度所不可或缺的最重要的基本资料<sup>[5, 7-8]</sup>。

随着探空技术的发展, 我国自2011年已经完成

由59-701型探空系统向L波段电子探空仪系统的全面更新换代, 新一代探空系统的显著优势是可以提供秒级别的、高垂直分辨率的探空数据<sup>[9]</sup>。近年来, 高垂直分辨率探空数据被广泛应用于各项科学研究中, 均证实了其独特的应用价值。例如, 杨湘婧等<sup>[10]</sup>利用垂直高分辨率探空数据建立模型, 进行大气边界层中近地面层的信息分析。卞建春等<sup>[11]</sup>利用垂直高分辨率探空数据对北京上空下平流层重力波的特性进行统计分析。Basha等<sup>[12]</sup>利用高垂直分辨率无线电探空的长期观测, 研究了热带站点Gadanki的大气边界层高度, 指出高垂直分辨率探空资料可以为边界层高度研究提供较优的资料基础。此外, 高分辨率探空数据还广泛应用于大气晴空湍流<sup>[13-14]</sup>、风切变等研究中<sup>[5, 13, 15]</sup>。

此外, 数值天气模拟, 资料同化及数值天气预报均可受益于高分辨率探空数据的获取和应用<sup>[16]</sup>, 探空

收稿日期: 2017年7月27日; 修回日期: 2017年12月10日

第一作者: 李庆雷(1987—), Email: liql@cma.gov.cn

通信作者: 周自江(1970—), Email: zzzj@cma.gov.cn

资助信息: 国家气象科技创新工程项目(CMAGGTD003-5);  
国家气象信息中心青年课题(NMICQJ201703)

观测系统越密，它对分析精确度的贡献越大，这个事实在水平和垂直方向都成立<sup>[17]</sup>。例如，郝民等<sup>[18]</sup>比较秒级探空与传统探空资料使用对资料同化预报带来的影响，指出高垂直分辨率探空资料的使用对数值模式分析初值及预报的改进有着积极的意义。姚爽等<sup>[7, 19]</sup>将L波段高分辨率探空数据用在WRF模式中进行变分同化试验，指出同化分钟数据可一定程度提升模式初始场品质并有助于改善对流层高空风和大量级降水预报性能。

然而，前人关于L波段秒级数据的同化应用研究<sup>[7, 18-19]</sup>主要基于高垂直分辨率探空资料在时间上的自然稀疏化，如利用时间分辨率较低的分钟探空数据。稀疏化是高频资料在同化应用过程中非常重要的一个技术环节<sup>[20]</sup>。而基于时间分辨率降低的自然稀疏化，如分钟探空数据，并不能保证每分钟都恰好是探空垂直廓线中气象要素的变化拐点，也就不能精准地解析出气象要素垂直变化的细致特征。在同化应用时，需要将探空廓线上的数据线性插值到模式面上<sup>[18]</sup>，应用自然稀疏化后的分钟数据就很容易造成较大的误差。

为解决上述问题，本文提出一套针对中国探空L波段秒级资料的、自适应的稀疏化技术方案，该方案可以实现从秒级探空资料中，提取温度、湿度、风向、风速等气象要素的垂直廓线的显著拐点，并整合各规定等压面层、对流层顶等重要探空层次，从而形成大约有一两百层的完整探空廓线。所得廓线的层数与欧美发达国家水平相当，并能更准确描绘各个气象要素垂直变化的精细结构特征。基于模式背景场（ERA-Interim再分析资料）的观测资料偏差评估技术，对比分析不同方法得到探空廓线上气象要素的数据量、偏差（Bias）和均方根误差（RMSE）等定量指标，对该稀疏化方案的效果进行了系统的检验。

## 1 资料和方法

### 1.1 研究资料

本文用到的L波段秒级探空数据来自国家气象信息中心发布的中国高空L波段秒级观测基础数据集（V1.0），该数据经过严格的质量控制，如允许值范围检查、僵值检查、垂直一致性检查、气候界限值检查、滤波检查、单调性检查、要素间一致性检查等等。由于秒级数据的质量控制方案不是本文研究重点，不赘述。值得一提的是，本文的稀疏化方案基于质量控制之后的秒级探空数据，以保证稀疏化后得到的数据信息的准确性。

## 1.2 研究方法

### 1.2.1 稀疏化方法

常规高空气象观测业务规范<sup>[21]</sup>中给出了选取温度特性层、湿度特性层、风特性层条件和步骤，例如，选取温度梯度的显著转折点，即两层间的温度分布与用直线连接的温度比较，大于1℃（第一个对流层顶以下）或大于2℃（第一个对流层顶以上）的差值最大的气层，作为选取的温度特性层；然后重复该方法直至相邻两层间没有特性层为止。在观测业务规范中，湿度特性层和风特性层的选择与上述步骤类似。

本文参考业务观测规范中特性层的选取条件<sup>[21]</sup>，针对探空高垂直分辨率资料进行稀疏化，根据气象要素垂直廓线的变化特征，自适应地从低到高逐层挑选出廓线上的显著转折点，即温度特性层、湿度特性层、风特性层、对流层顶等，以期用较少的数据点就能精准描绘气象要素的垂直变化特征。这里提取各个探空气象要素特性层的步骤简要介绍如下：（1）在一条完整的探空廓线上，选择地面层（探空数据第1秒）为第一个特性层 $L_1$ ；（2）以第1秒和第 $N$ 秒的数据点做直线，计算实际廓线到此直线的偏移量；（3）当此偏移量大于某一阈值时，即选定该层为第二个特性层 $L_2$ ；（4）然后以 $L_2$ 为基点，继续向上依次选出各个特性层 $L_3, L_4, \dots, L_n$ ；（5）值得一提的是，不同的气象要素在不同的探空高度，选取的阈值大小不同，这需要反复试验确定以满足不同的同化应用需求。

为便于实际应用的需要，该稀疏化方案还实现了直接从秒级探空数据中提取各个规定等压面的气象要素，与上面提取的特性层数据整合在一起，从而形成一条完整的大约有100~200层的探空廓线。

### 1.2.2 稀疏化效果检验方法

为了检验稀疏化效果，通过比较不同来源的探空观测数据和ERA-Interim再分析资料之间的偏差（Bias）和均方根误差（RMSE）来量化检验。其中偏差Bias代表一定时段内再分析资料平均值与探空实测资料平均值之差，而均方根误差

$$RMSE = \sqrt{\sum_{i=1}^N (ERA - OBS)^2 / N}, \quad (1)$$

是ERA-Interim再分析数据与探空实测数据偏差的平方和与观测次数比值的平方根。这些不同来源的探空廓线分别是：基于“中国高空特性层定时值数据集（V1.0）”和“中国高空规定等压面定时值数据集（V2.1）”整合的探空廓线数据（CMA-Merge），基于本文提出的稀疏化方案得到的探空廓线数据

(CMA-Thin), 以及美国国家环境预报中心研发的气候预测系统再分析资料(CFSR)中的探空廓线数据。

已有研究表明, 尽管由于受到数值预报模式和同化方案等系统性误差的影响, 再分析资料与观测结果相比可能仍然存在一定的差异<sup>[22]</sup>, 但是它们基本能合理反映东亚及其中国区域气候变化的时空分布特征<sup>[23-25]</sup>。因此, 文中应用ERA-Interim再分析资料作为背景场, 通过比较不同数据源与背景场的偏差(Bias)和均方根差(RMSE)来量化分析稀疏化效果。

## 2 结果与分析

### 2.1 稀疏化效果对比分析

单次探空廓线稀疏化效果如图1所示, 以53068站(二连浩特, GCOS探空站)2016年1月1日00 UTC时次的L波段探空温度垂直廓线为例, 可以看出, 原始的秒级观测数据有4318层, 中国高空特性层定时值数据集(V1.0)给出的温湿特性层数目是30层, 而应用本文设计的基于秒级资料的稀疏化方法得到的温湿特性层数目是118层。对比三条廓线可以很清楚地看出, 通过稀疏化方法得到的温度廓线, 较现有的30层的温度廓线, 通过捕捉更小尺度的温度拐点, 能更准确地描绘温度的垂直变化特征。值得一提的是, 当探空温度应用于数值模式资料同化过程中时, 需要将温度特性层的温度值线性插值到特定模式层<sup>[18-19]</sup>, 显然特性层层数多的温度廓线能够提供更精准的温度值。

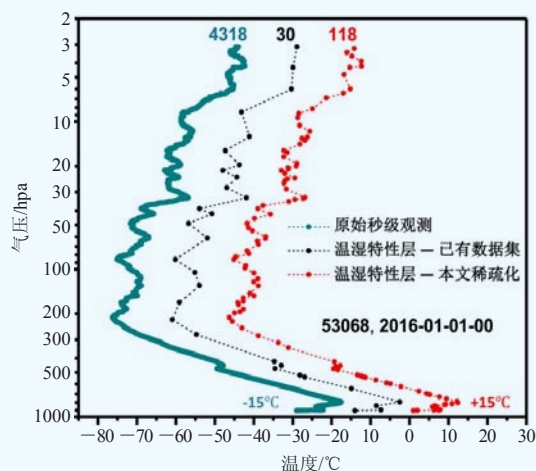


图1 利用两种方法得到的温度垂直廓线的对比图(为表示清楚, 图中曲线进行了平移)

Fig. 1 Comparison of temperature profiles between two methods (The curve is shifted in the figure for clarity)

### 2.2 温湿度特性层数目的时空分布

为了进一步检验稀疏化效果, 本文选取2014年全年中国120站00 UTC和12 UTC两个时次所有的秒级探空廓线数据进行统计分析, 得到各个站点平均单次探

空廓线提取得到的温湿特性层数目, 其不同取值范围的站点数分布如表1所示。图2给出了平均单次廓线中温湿特性层数目的空间分布图, 其中图2a统计的是已有的特性层数据集, 图2b则给出本文稀疏化得到的特性层数目的空间分布。首先, 通过表1和图2都显示, 已有的特性层数据集给出的温湿特性层数目平均值集中在20~30层, 而基于秒数据稀疏化得到的温湿特性层数目平均值绝大多数在60~90层, 显著地增加了特性层的数目。其次, 可以看出, 尽管两种方法给出的特性层数目差异很大, 但是图2a和2b给出的特性层数目的空间分布型表现出很好的一致性; 如, 两种方法统计得到的特性层数目相对较多的站点都集中分布在海拔较低的中国东南部。最后需要指出, 两图中用红色圆圈标记的7个探空站点, 是GCOS探空站, 由于施放大球, 其最大探测高度明显高于周边非GCOS探空台站, 用两种方法得到的特性层数目均多于周边站点。这些都说明, 本文提出的基于L波段秒级探空数据的稀疏化方案对于全国120个探空站都具有很好的普适性。

气象要素的垂直变化不仅与站点的地理分布有

表1 对比两种特性层提取方法得到的不同温湿特性层数目的站点数分布表

Table 1 Comparison of the station numbers between two methods

选取方法	特性层数取值范围	站点数
已有特性层数据集	[13, 22]	20
	[22, 23]	17
	[23, 24]	21
	[24, 25]	14
	[25, 27]	32
	[27, 30]	16
稀疏化方案	[39, 67]	19
	[67, 73]	19
	[73, 77]	20
	[77, 82]	20
	[82, 87]	21
	[87, 97]	21

表2 对比不同季节, 本文稀疏化方案得到不同温湿特性层数目的站点数分布表

Table 2 The station numbers in 4 seasons resulted from this sparsification scheme

特性层数取值范围	春	夏	秋	冬
[20, 50]	4	0	4	4
[50, 65]	8	6	66	11
[65, 80]	30	65	50	27
[80, 90]	26	44	0	33
[90, 100]	37	5	0	30
[100, 115]	15	0	0	14

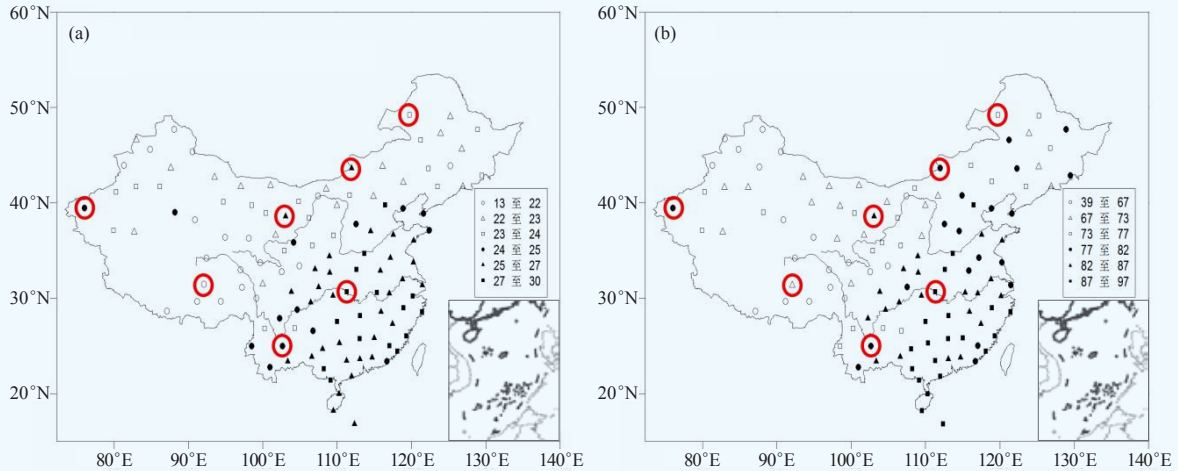


图2 两种方法得到平均单次廓线中温湿特性层数目的空间分布: (a) 已有特性层数据集; (b) 本文稀疏化  
Fig. 2 Distribution map of numbers of temperature and humidity significant levels in average single profile: (a) from the significant level datasets; (b) from the sparsification of high-resolution data

关, 而且在不同的季节同样应该表现出不同变化规律。表2给出了春、夏、秋、冬四个季节, 不同温湿特性层数目取值范围的站点数分布。如图3给出了稀疏化得到平均单次廓线中温湿特性层数目在不同季节

的空间分布结果。通过图表可以看出, 四个季节的特性层数目的空间分布型与图2中给出的全年分布图基本吻合。值得一提的是, 绝大多数的探空站点在冬春季节提取的特性层数目普遍多于夏秋季节, 这与夏秋

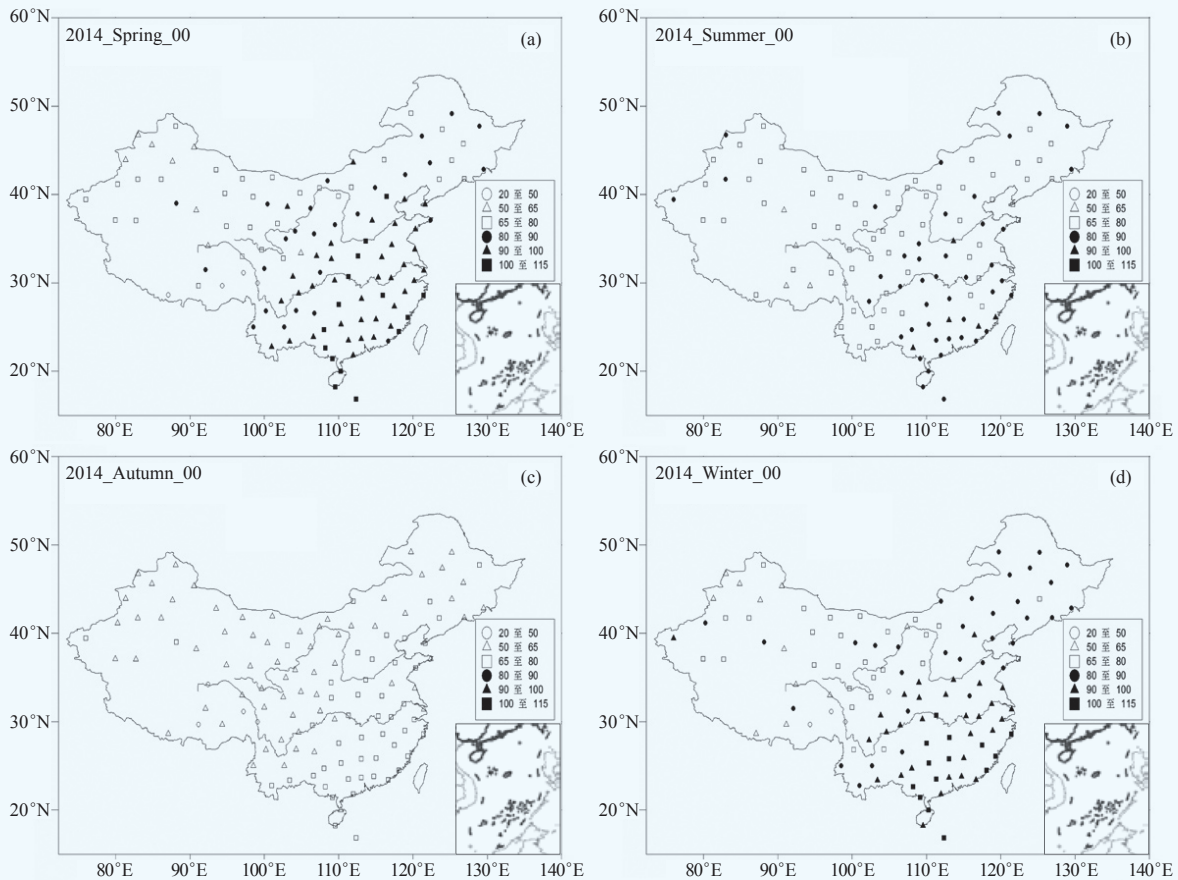


图3 本文稀疏化方案得到平均单次廓线中温湿特性层数目在不同季节的空间分布: (a) 春季; (b) 夏季; (c) 秋季; (d) 冬季

Fig. 3 Distribution maps of temperature and humidity significant level numbers in average single profile for: (a) spring; (b) summer; (c) autumn; (d) winter

季节整个大气层的温度较高导致其能量上下交换比较充分有关。

### 2.3 基于 ERA-Interim 再分析资料的稀疏化效果检验

首先，从时间序列分布上看，图4对比给出了两种数据源相对ERA-Interim再分析资料的检验效果，可以看出，基于秒级资料稀疏化方案得到的参与对比的探空温度数据量远远大于通过已有探空数据集整合的结果，而相对ERA-Interim再分析资料的温度偏差 (*Bias*) 和均方根误差 (*RMSE*) 的水平相当，并没有因为数据量的增大而增大；而且无论是*T*数据量，还是*T*偏差和*T*均方根误差，随时间变化的走势都表现出很好的一致性。

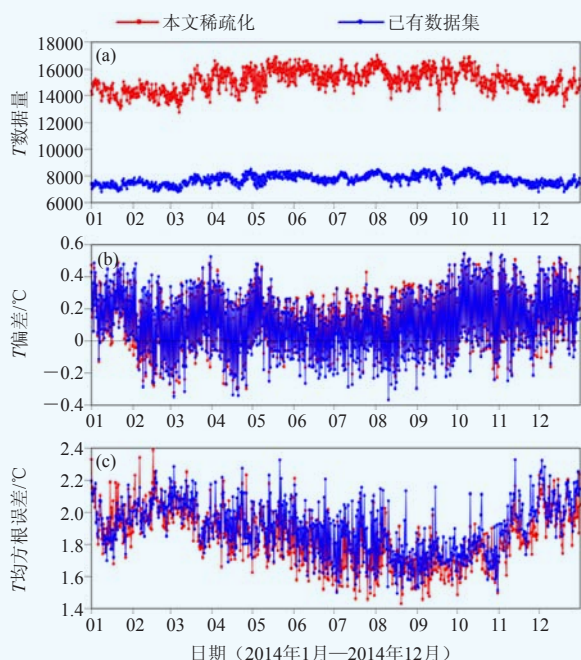


图4 两种数据源相对ERA-Interim再分析资料的评估结果：(a) 数据量；(b) 偏差 (*Bias*)；(c) 均方根误差 (*RMSE*)

Fig. 4 Evaluation results of two data sources compared with ERA-Interim reanalysis data: (a) data size; (b) *Bias*; (c) *RMSE*

其次，从检验结果的空间分布上看，图5对比给出了三种数据源的温度相对ERA-Interim再分析资料检验结果的垂直分布图，其中图5a代表在不同探空高度处参与对比分析的*T*数据量，图5b是温度偏差 (*Bias*)，图5c是温度均方根误差 (*RMSE*)。由图5a可以看出，由于CFSR中中国站点数目较少，使得在任一探空高度处的数据量均最少；而稀疏化得到的温度数据量又明显多于已有数据集整合的数据量。由图5b可以看出，经稀疏化得到的温度偏差较其他两个数

据源都更接近于0 °C，且其垂直变化幅度更小，较其他两个数据源表现出更好的垂直一致性。由图5c可以看出，在探空高度925~500 hPa，经稀疏化得到的温度均方根误差与整合数据源水平相当；而在探空低层1000~925 hPa和高层500~100 hPa，经稀疏化得到的温度均方根误差比整合数据源的要偏大，而这可能与参与统计的数据量较多有关。

总之，无论是从检验结果的时间序列分布 (图4)，还是其垂直空间分布 (图5)，经稀疏化方案得到的数据源都表现出明显的优势：其*T*数据量更大，*T*<sub>RMSE</sub>水平相当，*T*<sub>Bias</sub>更小。值得一提的是，不仅仅是温度*T*这一气象要素的稀疏化效果表现出上述统计特征，其他气象要素如湿度*Rh*、风速*V*等可以得到类似的结论，此处不赘述。

### 3 结论与讨论

本文基于L波段探空秒级资料设计的稀疏化方案，能够显著增加用于同化应用的探空廓线特性层数目，通过对120个探空站点2014年所有的秒级探空做稀疏化处理，得到的特性层数目的空间分布与通过特性层数据集得到的空间分布具有很好的一致性，而且其层数空间分布表现出明显的季节变化特征，这体现了本文的稀疏化方案的可行性与普适性。

基于ERA-Interim再分析资料做背景场，稀疏化后的探空数据较已有探空层数偏少的整合数据的偏差评估效果更好，通过*Bias*和*RMSE*的时间序列图以及垂直空间分布图可以清楚地对比三种方法得到探空数据的差异。随着数值预报模式的发展尤其是模式垂直分辨率的提高，必将促使更多层的探空数据进入同化应用。本文提出的稀疏化方案为高垂直分辨率探空资料的有效同化应用提供了保证。

此外，该稀疏化方案改进了提取特性层的算法，显著增加特性层层数，可以更好地应用于气候变化等研究中<sup>[6, 26-27]</sup>。在接下来的研究中，可以结合不同的数值预报模式，如GRAPES等，针对特定的天气过程进行个例分析，将稀疏化后的探空数据用于数值预报同化中，可以更好地体现该稀疏化方案的实际应用效果。

#### 参考文献

- [1] Ingleby B, Edwards D. Changes to radiosonde reports and their processing for numerical weather prediction. *Atmospheric Science Letters*, 2015, 16(1): 44-49.
- [2] 李岩瑛, 张强, 薛新玲, 等. 民勤大气边界层特征与沙尘天气的气候学关系研究. *中国沙漠*, 2011, 31(3): 757-764.
- [3] 唐家萍, 谭桂容, 谭畅. 基于L波段雷达探空资料的重庆市低空逆温特征分析. *气象科技*, 2012, 40(5): 789-793.
- [4] Bodeker G E, Bojinski S, Cimini D, et al. Reference upper-air observations for climate: From concept to reality. *Bulletin of the*

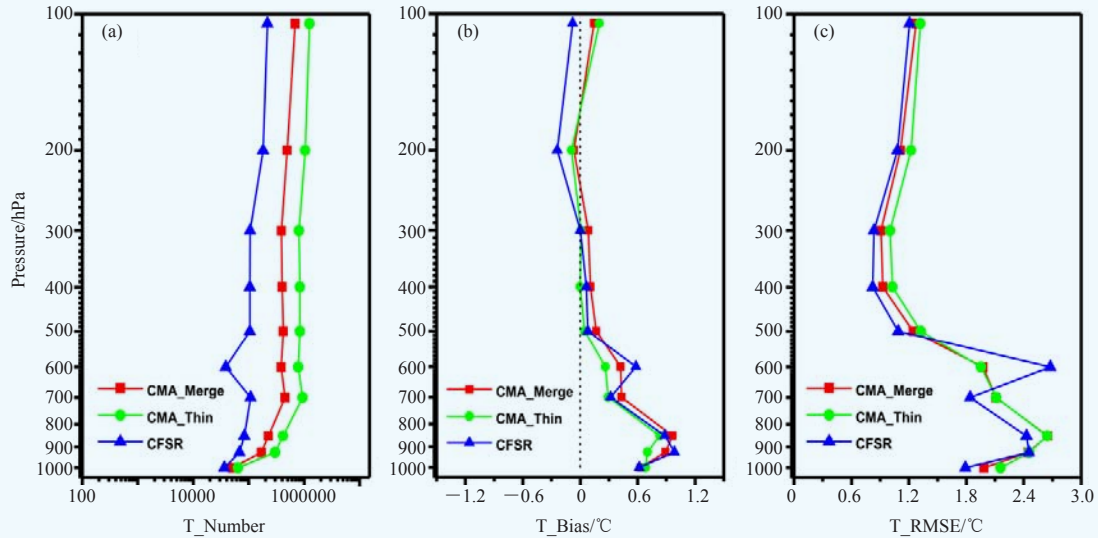


图5 三种数据源（整合数据源CMA-Merge，稀疏化数据源CMA-Thin，CFSR）相对于ERA-Interim再分析资料的评估结果的垂直分布图：（a）数据量；（b）偏差（Bias）；（c）均方根误差（RMSE）

Fig. 5 Vertical distribution of evaluation results of the ERA-Interim reanalysis data from three data sources (the CMA-Merge, CMA-Thin and the CFSR): (a) Data size; (b) Bias; (c) RMSE

- American Meteorological Society, 2016, 97(1): 123-135.
- [5] Ingleby B, Pauley P, Kats A, et al. Progress toward high-resolution, real-time radiosonde reports. *Bulletin of the American Meteorological Society*, 2016, 97(11): 2149-2161.
- [6] 阮新, 朱艳峰, 鞠晓慧. 融合特性层观测资料推算探空站海拔高度的方法. *高原气象*, 2011, 30(2): 532-537.
- [7] 姚爽. L波段探空数据分析及在区域模式中的同化应用试验. 中国气象科学研究院硕士学位论文, 2014.
- [8] Durre I, Yin X. Enhanced radiosonde data for studies of vertical structure. *Bulletin of the American Meteorological Society*, 2008, 89(9): 1257-1262.
- [9] 李伟, 刘凤琴, 徐磊, 等. L波段高空气象探测系统软件. *气象科技*, 2008, 36(2): 237-239.
- [10] 杨湘婧, 徐祥德, 陈宏尧, 等. L波段探空高分辨率廓线中近地层信息分析及相关模型. *气象*, 2011, 37(12): 1504-1510.
- [11] 卞建春, 陈洪滨, 吕达仁. 用垂直高分辨率探空资料分析北京上空下平流层重力波的统计特性. *中国科学 D 辑: 地球科学*, 2004, 24(8): 748-756.
- [12] Basha G, Ratnam M V. Identification of atmospheric boundary layer height over a tropical station using high-resolution radiosonde refractivity profiles: Comparison with GPS radio occultation measurements. *Journal of Geophysical Research Atmospheres*, 2009, 114(11): 713-721.
- [13] Love P T, Geller M A. Research using high (and higher) resolution radiosonde data. *Eos Transactions American Geophysical Union*, 2012, 93(35): 337-338.
- [14] Clayton C A, Kantha L. On Turbulence and mixing in the free atmosphere inferred from high-resolution soundings. *Journal of Atmospheric and Oceanic Technology*, 2008, 25(6): 833-852.
- [15] Leena P P, Venkat Ratnam M, Krishna Murthy B V, et al. Detection of high frequency gravity waves using high resolution radiosonde observations. *Journal of Atmospheric and Solar-Terrestrial Physics*, 2012, 77(3): 254-259.
- [16] Hamilton K, Vincent R A. High-resolution radiosonde data offer new prospects for research. *Eos, Transactions American Geophysical Union*, 1995, 76(49): 497-497.
- [17] Kalnay E. 大气模式、资料同化和可预报性. 北京: 气象出版社, 2005.
- [18] 郝民, 田伟红, 龚建东. L波段秒级探空资料在GRAPES同化系统中的应用研究. *气象*, 2014, 40(2): 158-165.
- [19] 姚爽, 陈敏, 王建捷. L波段分钟数据在WRF模式中的变分同化应用试验. *气象*, 2015, 41(6): 695-706.
- [20] Data usage and quality control for ERA-40, ERA-interim and the operational ECMWF data assimilation system. ERA report series, 2011.
- [21] 中国气象局. 常规高空气象观测业务规范. 北京: 气象出版社, 2010.
- [22] 赵天保, 符淙斌, 柯宗建, 等. 全球大气再分析资料的研究现状与进展. *地球科学进展*, 2010, 25(3): 242-254.
- [23] 赵天保, 符淙斌. 应用探空观测资料评估几类再分析资料在中国区域的适用性. *大气科学*, 2009, 33(3): 634-648.
- [24] 支星, 徐海明. 三种再分析资料的高空温度与中国探空温度资料的对比分析. *高原气象*, 2013, 32(1): 97-109.
- [25] 赵佳莹, 徐海明. 中国区域位势高度场探空与再分析资料的对比分析. *气象科学*, 2014, 34(2): 128-138.
- [26] 阮新, 熊安元, 胡开喜, 等. 中国历史探空资料部分等压面位势高度订正. *应用气象学报*, 2015(03): 257-267.
- [27] 郭艳君, 张思齐, 颜京辉, 等. 中国探空观测与多套再分析资料气温序列的对比研究. *气象学报*, 2016, 74(2): 271-284.