0577-6619/2011/69(5)-0872-82 *Acta Meteorologica Sinica* 气象学报

亚高斯贝叶斯预报处理器及其初步试验*

陈法敬[1] 矫梅燕[2] 陈 静[3]
CHEN Fajing[1] JIAO Meiyan[2] CHEN Jing[3]

1. 南京信息工程大学大气科学学院,南京,210044
2. 中国气象局,北京,100081
3. 国家气象中心,北京,100081
1. *College of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing* 210044, *China*
2. *China Meteorological Administration, Beijing* 100081, *China*
3. *National Meteorological Center, Beijing* 100081, *China*
2009-12-15 收稿,2010-09-25 改回.

Chen Fajing, Jiao Meiyan, Chen Jing. 2011. The meta-Gaussian Bayesian Processor of Forecast and its preliminary experiments. *Acta Meteorologica Sinica*, 69(5):872-882

Abstract One of the trend of the public weather service is to provide users with probabilistic weather forecasts. The continuous improvement of probabilistic forecasting techniques realizes the constant optimization of probabilistic forecast information. Among numerous techniques and methods for probabilistic forecasting, the Bayesian Processor of Forecast (BPF) is a new statistical probabilistic forecasting technique based on the Bayesian statistical theory. The BPF can, based on the Bayesian statisical theory, transform a deterministic forecast from a deterministic forecasting system into a probabilistic forecast according to the statistical relationship between historical observations and forecasts from that system that is able to denote the forecasting performance of that deterministic system to quantify the forecasting uncertainty of that deterministic forecast. The meta-Gaussian likelihood model is suitable for several kinds of stochastic dependence structures with monotone likelihood ratio, so the meta-Gaussian BPF adopting this kind of likelihood model can be flexibly applied in many fields, such as meteorology, and hydrology. After the Bayes theorem with two continuous random variables and the normal-linear BPF are briefly introduced, this paper discusses the meta-Gaussian BPF for a continuous predictand using one single predictor. In order to test its performance, a preliminary experiment of the meta-Gaussian BPF is carried out, using daily control forecasts (with a leadtime of 96 hours) from the NMC, ECMWF and NCEP ensemble predictions of surface temperature (T_{2m}) at 00:00 UTC at Changsha station and Wuhan station during January 2008 as the deterministic forecasting data. The analysis of experiment results shows that the meta-Gaussian BPF can transform a control forecast of T_{2m} from any one of the three ensemble predictions into a probabilistic forecast of T_{2m}, which quantifies the forecasting uncertainty of that control forecast; the performances of the T_{2m} probabilistic forecasts obtained from different control forecasts are different from each other.

Key words Meta-Gaussian likelihood model, BPF, Forecasting uncertainty, Probabilistic forecasting

摘 要 为用户提供概率天气预报信息是公共气象服务的发展趋势,概率天气预报技术的不断改进实现了概率天气预报信息的不断优化。在众多概率天气预报技术方法中,贝叶斯预报处理器是一种新近出现的、基于贝叶斯统计理论的概率预报技术;贝叶斯预报处理器可以根据一个确定性预报系统的预报值与观测值之间代表着这个系统预报性能的统计关系,借助于贝叶斯统计理论,把一个确定性预报转化为一个概率预报,从而实现对预报不确定性的定量化。由于亚高斯似然模型可以适用于多种单调似然比随机依赖结构,故采用该似然模型的亚高斯贝叶斯预报处理器,它在气象、水文等领域具有较强的适用性。

* 资助课题:国家公益性行业(气象)科研专项(GYHY200706001、GYHY200906007)、国家自然科学基金面上项目(41075035)。
作者简介:陈法敬,主要从事数值预报产品释用研究。E-mail: chenfj@cma.gov.cn

在简要介绍了连续型二维随机变量情形下的贝叶斯定理及正态-线性贝叶斯预报处理器之后,详细论述了采用单一预报因子的连续型预报量亚高斯贝叶斯预报处理器,并以长沙站和武汉站2008年1月每日00时(世界时)地面气温(T_{2m})的中国国家气象中心、欧洲中期天气预报中心、美国国家环境预测中心集合预报中的控制预报资料(预报时效选为96 h)作为确定性预报样本,对亚高斯贝叶斯预报处理器进行了初步试验。结果表明,亚高斯贝叶斯预报处理器可以将T_{2m}各集合预报中的控制预报转化为能定量地表达各控制预报不确定性的T_{2m}概率预报;源自不同控制预报的亚高斯贝叶斯预报处理器T_{2m}概率预报的性能存在差异。

关键词 亚高斯似然模型,贝叶斯预报处理器,预报不确定性,概率预报

中图法分类号 P456.8 P457.3

1 引 言

确定性预报向概率预报转变是天气预报的发展趋势。对于某一天气要素,一个确定性预报系统(可能是某一个数值预报模式或天气预报机构)会给出一个确定性的预报产品。实际上,这个确定性预报也是它的制作者不自觉地制作的一种"概率预报":预报量在确定性预报值处出现的概率为1,在其他值处为0。实际上,实况并不会与确定性预报值完全重合,而是会或近或远地落在它的左右。这样,确定性预报就否定了实况出现在确定性预报值之外的可能性,即屏蔽了客观存在的预报不确定性,据此而提供的预报信息是不完善的(周秀骥,2005)。概率天气预报的任务,是要为其使用者提供预报制作者对天气状况未来发展趋势认知状态的较完整表达;其中,很重要的一点就是对预报不确定性的定量化表达。因此,相对于确定性预报,概率预报能够提供更完善的预报信息(俞小鼎等,2004)。对于一个确定性预报系统,需要利用合适的概率预报技术,以实现其向概率预报系统的角色转换,即对确定性预报不能表达出来的预报不确定性进行量化,从而得到包含有较完善预报信息的概率预报。

随着天气预报技术的不断发展(张大林,2005;矫梅燕等,2006;薛纪善,2006),已提出多种概率预报技术(张军等,1998;黄嘉佑,2000),集合预报就是其中一种(范新岗,1999;刘金达,2000;李泽椿等,2002;陈静等,2002;陈超辉等,2010)。在概率预报统计技术中,现在较常用的是模式输出统计(MOS)技术。在运用MOS分类概率预报技术时,需要将众多的模式输出预报因子进行一种称为格点-二分类转换的处理,这种转换过程会丢失预报因子的一部分预报信息,获得的结果是分类预报量离散的概率预报值,可能会导致获取预报量在一些极端值(或极端区域)上的概率预报值时遇到困难(Glahn, et al,1972;Antolik,2000;Krzysztofowicz, et al,2006b)。

基于贝叶斯统计理论的概率预报技术方法,中国水文领域已进行了研究与引进介绍(王善序,2001;张洪刚等,2004,2006),但在气象领域,这方面的研究及引进介绍较少(马培迎,1999;陈朝平等,2010;陈法敬等,2011)。在其他国家,Krzysztofowicz等(1983,1986,1987,1991)曾使用过正态-线性贝叶斯预报处理器(BPF),但所采用的正态-线性似然模型使其应用仅限于少数近似服从正态分布的预报变量,以及预报变量间多为线性、同方差随机依赖结构的短期预报。Krzysztofowicz(2002)在正态-线性BPF的基础上,引入正态分位数转换(NQT),开发出了亚高斯贝叶斯预报处理器(亚高斯BPF)。由于采用了亚高斯似然模型,亚高斯BPF可应用于很多不服从正态分布的预报变量及预报变量间多为非线性、异方差随机依赖结构的中长期预报,在气象、水文等领域正逐渐得到应用(Kelly, et al,1995,1997;Krzysztofowicz, et al,2000a,2000b;Herr, et al,2005)。

亚高斯BPF具有3种形式,对应于3种类型的预报量:二分类预报量(如降水发生指标,即有或无降水)、多分类预报量(如降水类型指标,即降水的雨、雪等形式)、连续型预报量(如温度、能见度、风速及以降水发生为条件的降水量等)。在亚高斯BPF中,预报量可以有多个预报因子,共同构成一个预报因子向量;通常,预报因子向量包含1—3个预报因子(Krzysztofowicz, et al, 2006a, 2006b)。基于Krzysztofowicz等的研究,本文旨在引进亚高斯BPF概率预报技术。在介绍连续型二维随机变量情形下的贝叶斯定理及正态-线性BPF、亚高斯BPF之后,将基于单一预报因子的连续型预报量亚高斯BPF,利用2008年1月中国国家气象中心(NMC)、欧洲中期天气预报中心(ECMWF)、美国国家环境

预测中心(NCEP)集合预报系统控制预报的逐日 00 时(世界时,下同)的地面气温(T_{2m})96 h 的预报资料,结合长沙和武汉两站观测资料,对亚高斯 BPF 进行初步试验,并对源自不同控制预报的亚高斯 BPF 概率预报能力进行检验评估。

2 贝叶斯预报处理器及正态-线性贝叶斯预报处理器

2.1 贝叶斯预报处理器

一个连续型预报量 y 及其一个连续型预报因子 x 共同构成一个连续型二维预报变量(x,y)。记(x,y)的联合概率密度函数为 $h(x,y)$,预报变量(x,y)的边缘概率密度函数分别为 $k(x)$、$g(y)$,边缘概率分布函数分别记为 $K(x)$、$G(y)$,条件概率密度函数分别记为 $f(x|y)$、$\phi(y|x)$;预报量 y 的似然密度函数也记为 $f(x|y)$。若已知预报因子的值 x,则条件概率密度函数 $\phi(y|x)$也称为 y 的后验概率密度函数,相应地,$g(y)$也称为 y 的先验概率密度函数。

根据概率论中的贝叶斯定理,二维预报变量(x,y)的概率密度函数间存在关系式

$$\phi(y \mid x) = \frac{h(x,y)}{k(x)} = \frac{f(x \mid y)g(y)}{k(x)} \tag{1}$$

其中,$k(x) = \int_{-\infty}^{\infty} f(x \mid \zeta)g(\zeta)\mathrm{d}\zeta$。式(1)为贝叶斯预报处理器(BPF)概率预报公式的原始形式。若预报量 y 的先验概率密度函数 $g(y)$、似然密度函数 $f(x|y)$已确定,则由预报因子的一个预报值 x,利用式(1)就可以得到 y 的后验概率密度函数形式的概率预报 $\phi(y|x)$。

实际上,BPF 是一种在已知预报因子的预报值情况下,获得预报量后验概率分布的统计技术(Kelly, et al, 1995)。BPF 流程如图 1(依据 Krzysztofowicz, et al, 2006b)所示,利用预报量气候资料,以及某个确定性预报系统对该预报量的预报资料及对应的实况资料,建立起对应于这个确定性预报系统的 BPF 概率预报公式,藉此就可以将一个确定性预报转化为一个概率预报。

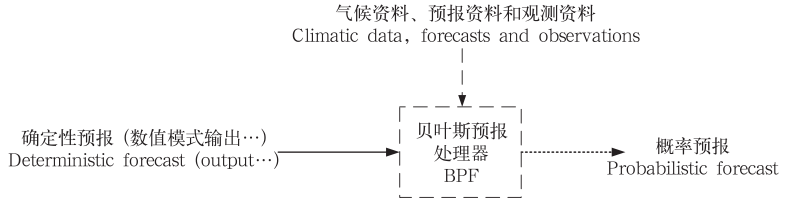


图 1 BPF 概念流程

Fig. 1 Conceptual flow chart for the BPF

2.2 正态-线性贝叶斯预报处理器

要建立 BPF,就需要为似然密度函数 $f(x|y)$确定一个合适的数学模型,即为 y 寻找似然模型。实际上,由于预报变量间存在复杂多样的随机依赖结构,y 的似然模型也是多种多样的。正态-线性 BPF 的似然模型具有正态-线性的数学形式。

在正态-线性 BPF 中,二维预报变量(x,y)服从二元正态分布,预报因子 x 服从正态分布 $N(\mu_x, \sigma_x^2)$,预报量 y 服从正态分布 $N(\mu_y, \sigma_y^2)$,其中,μ_x、σ_x^2 与 μ_y、σ_y^2 分别为 x 与 y 的均值、方差,协方差记为 σ_{xy},相关系数记为 ρ。预报变量 x、y 间的随机依赖结构具有线性回归形式

$$x = ay + b + \Theta \tag{2}$$

Θ 服从正态分布 $N(0,\sigma^2)$,且

$$E(x \mid y) = ay + b$$

$$\mathrm{Var}(x \mid y) = \sigma^2$$

线性回归系数 $a = \frac{\sigma_{xy}}{\sigma_y^2}$,$b = \mu_x - \frac{\sigma_{xy}}{\sigma_y^2}\mu_y$,$\sigma^2 = \sigma_x^2 - \frac{\sigma_{xy}^2}{\sigma_y^2}$;预报量 y 的似然模型为

$$f(x \mid y) = \frac{1}{\sigma} q\left(\frac{x - ay - b}{\sigma}\right) \tag{3}$$

其中,$q(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ 为标准正态分布的概率密度函数。

根据式(1)—(3),正态-线性 BPF 的概率预报公式(后验概率密度函数形式)为

$$\phi(y \mid x) = \frac{1}{\sigma_y \sqrt{1-\rho^2}} q\left(\frac{y - \left(\mu_y + \rho\sigma_y \frac{x - \mu_x}{\sigma_x}\right)}{\sigma_y \sqrt{1-\rho^2}}\right) \tag{4}$$

由式(4)可知,对预报量 y 的先验概率密度函数

$\frac{1}{\sigma_y}q(\frac{y-\mu_y}{\sigma_y})$ 进行修订后,得到的后验概率密度函数还是正态的,但均值、方差均有改变,其中,方差总是减小的。

正态-线性 BPF 的似然模型要求预报变量 x、y 间具有二元正态分布随机依赖结构,但在气象、水文等领域,符合这种要求的预报变量为数不多。因此,其适用范围较小。

3 亚高斯贝叶斯预报处理器

实际上,由于预报变量间的随机依赖结构是复杂多样的,因此,不能用式(1)的某一个固定似然模型(如正态-线性似然模型)来较好地统一刻画多种类型的随机依赖结构。为此,Kelly 等(1995)、Krzysztofowicz(1997, 2008)和 Krzysztofowicz 等(2006a, 2006b)在正态-线性似然模型的基础上,引入正态分位数转换(NQT),得到了亚高斯似然模型,从而开发出了具有较强适用性的亚高斯 BPF。

3.1 正态分位数转换

对于二维预报变量(x,y),正态分位数转换可表示为

$$\begin{cases} z = Q^{-1}(K(x)) \\ v = Q^{-1}(G(y)) \end{cases} \tag{5}$$

其中,Q^{-1}为标准正态分布 $Q(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\zeta^2/2}\mathrm{d}\zeta$ 的逆,$K(x)$、z 与 $G(y)$、v 分别为预报变量 x 与 y 的概率分布函数、转换预报变量;转换预报变量 z、v 构成二维转换预报变量(z,v)。

具有一定随机依赖结构的二维预报变量(x,y),经正态分位数转换后,其随机依赖结构将会发生什么样的变化呢? Kelly 等(1995)根据经验总结出 3 种类型的随机依赖结构并对其进行了统一表述:在二维预报变量(x,y)的概率密度空间中,若对于预报量 y 的任一对值$\{(y,y')\mid y<y'\}$,y 似然密度函数的比值$\frac{f(x\mid y')}{f(x\mid y)}$都是 x 的不减(增)函数,则称二维预报变量(x,y)具有正(负)似然比随机依赖结构;正、负似然比随机依赖结构统称为单调似然比随机依赖结构(Lehmann, 1966)。

若二维预报变量(x,y)具有单调似然比随机依赖结构,则在转换预报变量的概率密度空间中,二维转换预报变量(z,v)服从二元正态分布,即具有线性回归形式的随机依赖结构(Kelly, et al, 1995)。正态分位数转换的这个关于单调似然比随机依赖结构的性质是整个亚高斯似然模型的核心。

3.2 亚高斯贝叶斯预报处理器

3.2.1 技术思路

如上述,多数预报变量间都具有复杂多样的随机依赖结构,要想利用其制作概率预报,再沿用旧的似然模型建模思路,即在预报变量的概率密度空间中,寻找预报量的某一固定的既能较真实地刻画预报变量间的随机依赖结构,又具有较强适用性的似然模型是行不通的。Kelly 等(1995,1997)、Krzysztofowicz 等(2000a, 2000b)在正态-线性似然模型的基础上,引入正态分位数转换进行预报变量概率密度空间的转换,通过在转换预报变量概率密度空间中,建立转换预报变量随机依赖结构的似然模型,实现了对具有单调似然比性质的多种类型的随机依赖结构的统一刻画,从而为贝叶斯概率预报的似然模型建模提供了一条新途径。

若二维预报变量(x,y)具有单调似然比随机依赖结构(图 2a),则经过正态分位数转换后,在转换预报变量概率密度空间中,二维转换预报变量(z,v)服从二元正态分布,随机依赖结构具有线性回归形式(图 2b)。

在转换预报变量的概率密度空间 $o\text{-}zvh$ 中,对二维转换预报变量(z,v)运用正态-线性 BPF,得到转换预报量 v 的概率预报公式;之后,再通过正态分位数转换的逆转换及雅克比式,将 v 的概率预报公式转换到原始预报变量的概率密度空间 $o\text{-}xyh$ 中,就得到了原始预报量 y 的概率预报公式。这就是亚高斯 BPF 的技术思路。这样,如果说在利用亚高斯似然模型求取原始预报量 y 的概率预报公式的过程中,正态-线性 BPF 起到了“桥梁”的作用,则正态分位数转换的角色就是“引桥”;由此也可以看出,亚高斯 BPF 是正态-线性 BPF 的发展。

3.2.2 概率预报公式

现实中,预报因子 x 的可用历史预报资料通常要远短于预报量 y 的可用历史观测资料。因此,x 的边缘概率分布函数 $K(x)$以样本估计 $\overline{K}(x)$表达,而取 y 的边缘概率分布函数 $G(y)$为其气候概率分布。

将具有单调似然比随机依赖结构的二维预报变

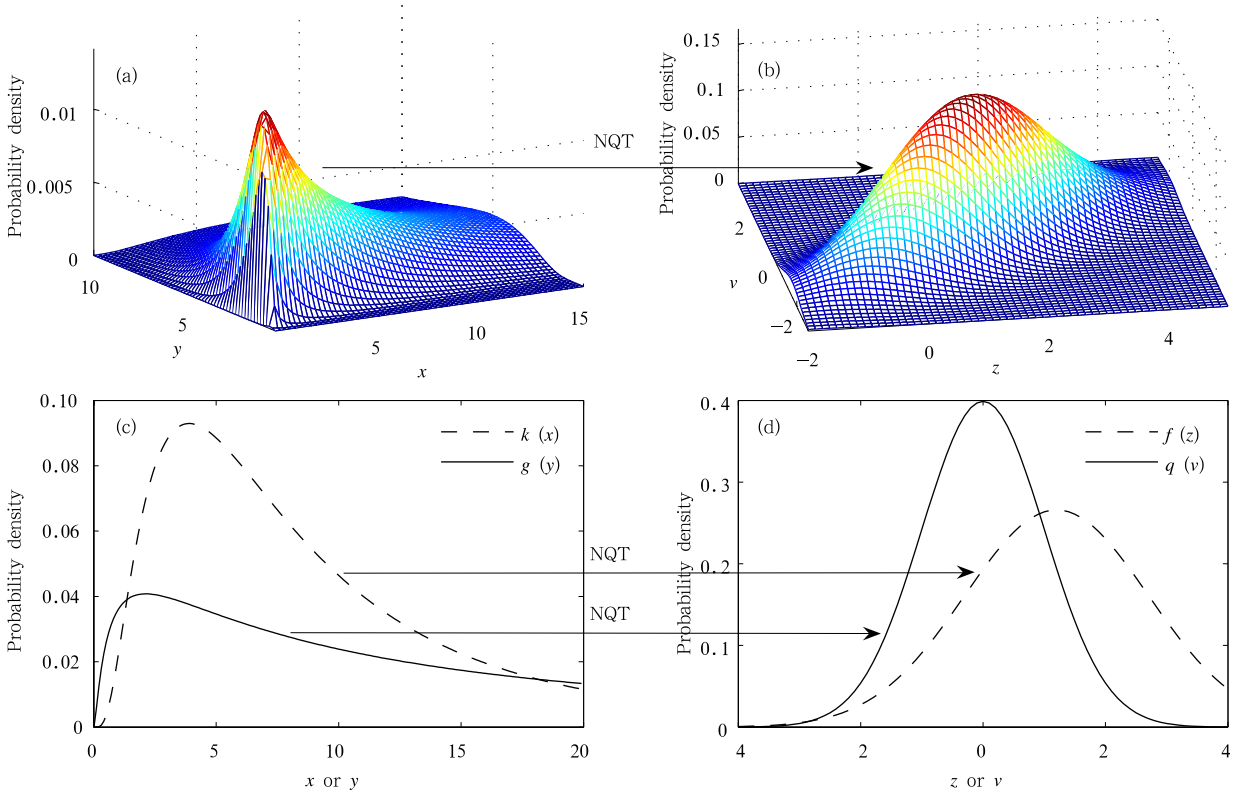


图 2 正态分位数转换示意图(a、b. 原始和转换概率密度,c、d. 原始和转换边缘概率密度)

Fig. 2 Sketch map of the NQT

(a, b. the original and the transformed probability density; c, d. the original and the transformed marginal probability density)

量(x,y)进行正态分位数转换转换(以$\overline{K}(x)$代替$K(x)$),得到服从二元正态分布的二维转换预报变量(z,v)。由正态分位数转换的性质,转换预报量v服从标准正态分布,而假定转换预报因子z服从正态分布(Kelly, et al, 1995)。二维转换预报变量(z,v)具有线性回归形式的随机依赖结构。由此,在转换预报变量概率密度空间o-zvh中,利用正态-线性BPF,可得转换预报量v的概率预报公式

$$\phi'(v \mid z) = \frac{\sqrt{a^2+\sigma^2}}{\sigma} \cdot q\left(\frac{v-a(z-b)/(a^2+\sigma^2)}{\sigma/\sqrt{a^2+\sigma^2}}\right) \tag{6}$$

需要将式(6)转换到原始预报变量概率密度空间o-xyh中,以得到原始预报量y的概率预报公式。正态分位数转换的雅克比式为

$$J(s) = \left|\frac{\mathrm{d}Q^{-1}(R(s))}{\mathrm{d}s}\right| = \frac{r(s)}{q(Q^{-1}(R(s)))} \tag{7}$$

式中,s代表预报变量x或y,R代表$\overline{K}$或G,r代表$\overline{k}$或g。由式(5)、(6)、(7),可得到原始预报量y的概率预报公式(概率密度函数形式)

$$\phi(y \mid x) = \frac{\sqrt{a^2+\sigma^2}\,g(y)}{\sigma q(Q^{-1}(G(y)))} \cdot q\left(\frac{Q^{-1}(G(y)) - a(Q^{-1}(\overline{K}(x)) - b)/(a^2+\sigma^2)}{\sigma/\sqrt{a^2+\sigma^2}}\right) \tag{8}$$

依据式(8),还可得到y另外两种形式的概率预报公式。概率分布函数形式

$$\Phi(y \mid x) = \int_{-\infty}^{y} \phi(\zeta \mid x)\mathrm{d}\zeta = Q\left\{\frac{1}{t}[Q^{-1}(G(y)) - c_1 Q^{-1}(\overline{K}(x)) - c_0]\right\} \tag{9}$$

以及后验概率分位数函数形式

$$y_{p|x} = G^{-1}(Q(c_1 Q^{-1}(\overline{K}(x)) + c_0 + tQ^{-1}(p))) \tag{10}$$

其中,$c_1 = \frac{a}{a^2+\sigma^2}$,$c_0 = \frac{-ab}{a^2+\sigma^2}$,$t = \left(\frac{\sigma^2}{a^2+\sigma^2}\right)^{1/2}$。

由此,y的概率密度函数形式的概率预报式(8)还可化为

$$\phi(y \mid x) = \frac{1}{t}\exp\left(\frac{1}{2}\{[Q^{-1}(G(y))]^2 - [Q^{-1}(\Phi(y \mid x))]^2\}\right) g(y) \tag{11}$$

式(9)、(10)、(11)即为亚高斯 BPF 三种形式的概率预报公式(以上公式推导依据 Kelly, et al, 1995; Krzysztofowicz, et al, 2006b; Krzysztofowicz, 2008)。

3.2.3 有效信息评分

在亚高斯 BPF 中,预报因子对预报量的预报价值,即预报因子预报能力的强弱用有效信息评分(I_S)来表示,也就是由 Krzysztofowicz(1992)提出的贝叶斯相关评分,I_S 计算式为

$$I_S = \left[\left(\frac{a}{\sigma}\right)^{-2} + 1\right]^{-\frac{1}{2}} \tag{12}$$

由式(12)可知,I_S 的值取决于由二维转换预报变量的线性回归系数计算得到的信噪比$|a|/\sigma$;回归斜率的绝对值$|a|$是信息指标,残差的标准差 σ 是噪音指标。I_S 的值域为(0, 1),且 I_S 值越大,则预报因子的预报能力越强(Krzysztofowicz, 2008)。

3.2.4 亚高斯 BPF 的特点

不同于正态-线性 BPF 采用的具有固定边缘概率分布的正态-线性似然模型,亚高斯 BPF 采用的亚高斯似然模型的原始预报变量的边缘概率分布不是固定的,其处理的单调似然比随机依赖结构形式也非唯一。亚高斯 BPF 具有区别于其他概率预报统计技术的性质:(1)不限制原始预报变量的边缘概率分布 $\overline{K}(x)$、$G(y)$,边缘分布可以是正态的或非正态的,可以是参数化的或非参数化的;(2)要求预报变量具有单调似然比随机依赖结构,但其可以是非线性(条件均值)的、异方差(条件方差)的,这两点使亚高斯 BPF 在实际应用中具有很强的灵活性及较广的适用范围;(3)概率预报公式是具有分析表达式的连续函数,因 Q 与 Q^{-1} 有多项式近似表达式,q 具有闭合形式,从而使亚高斯 BPF 便于数值计算(Krzysztofowicz, et al, 2006b)。

4 初步亚高斯 BPF 概率预报试验

4.1 试验要素

此次亚高斯 BPF 概率预报初步试验,对预报变量作如下选取:

预报量 T_k:长沙站或武汉站 1 年中第 k 天(1 月 1 日,$k=1$,闰年 2 月 29 日略)00 时气温(T_{2m});预报因子 X_k:NMC、ECMWF 或 NCEP 集合预报中的控制预报对长沙站或武汉站第 k 天 00 时 T_{2m} 的 96 h 预报。

预报量 T_k 及预报因子 X_k 选取的历史观测、预报资料为:T_k 的观测样本:长沙站(1987 年 1 月 1 日—2008 年 9 月 30 日),武汉站(1980 年 1 月 1 日—2008 年 9 月 30 日)每日 00 时 T_{2m} 观测值。X_k 的预报样本:2008 年 1 月 5 日—2 月 4 日,上述各集合预报控制预报对长沙站和武汉站每日 00 时 T_{2m} 的96 h预报资料。其中,1 月 5—31 日的样本,用于建立亚高斯 BPF 概率预报公式;2 月 1—4 日的样本用于独立概率预报试验;概率预报检验使用 1 月 5 日—2 月 4 日的所有样本。

从长沙站 T_k 在 1 年中的逐日气候均值 m_k、方差、标准差 σ_k(图 3)可以看出,T_k 的这些统计量具有明显的年周期性,这对预报因子 X_k 也是成立的。而亚高斯 BPF 所处理的预报变量须具有统计平稳性,为此,需要对预报变量做标准化处理

$$\begin{cases} T = \dfrac{T_k - m_k}{\sigma_k} \\ X = \dfrac{X_k - m_k}{\sigma_k} \end{cases} \tag{13}$$

标准化预报变量 T、X 分别称为标准化预报量、标准化预报因子。图4为2007年的长沙站 T_k 及

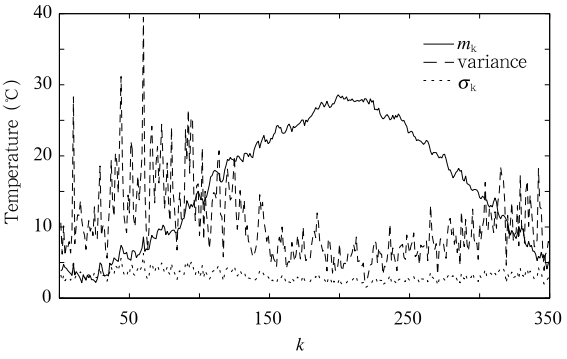


图 3 T_k 的逐日气候统计特征
(1987—2008 年,长沙站)
Fig. 3 Daily climatic statistical features of T_k at Changsha station (1987 - 2008)

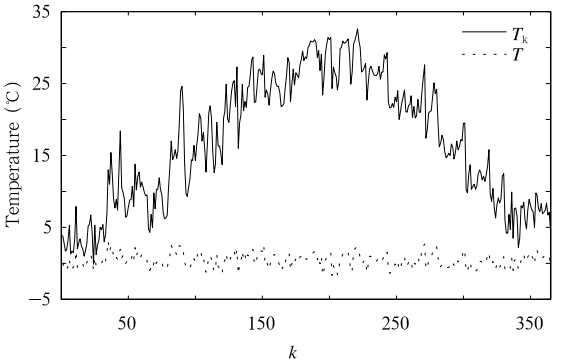


图 4 预报量 T_k 及标准化预报量 T(2007 年,长沙站)
Fig. 4 Values of T_k and the normalized T at Changsha station in 2007

T。可以看出,标准化预报量 T 的平稳性较好。

4.2 试验公式

根据气候统计学知识,T_k 的定时观测值近似服从正态分布(幺枕生等,1990):$T_k \sim N(\mathrm{m}_k, \sigma_k^2)$。由此,标准化预报量 T 服从标准正态分布,即取 $G(T)$ 为 $Q(T)$。因标准化过程中所用位移参数 m_k、尺度参数 σ_k 并非预报因子 X_k 自身的统计量,同时也为试验方便,取标准化预报因子 X 的边缘概率分布估计 $\overline{K}(x)$ 为正态分布 $N(\mu_x, \sigma_x^2)$,μ_x、σ_x 为 X 的均值、标准差。

至此,就可以利用预报量 T_k 和预报因子 X_k 的试验样本、T_k 的先验概率分布 $Q(\dfrac{T_k - m_k}{\sigma_k})$,以及 X_k 的边缘概率分布估计 $Q(\dfrac{\dfrac{x_k - \mathrm{m}_k}{\sigma_k} - \mu_x}{\sigma_x})$,针对不同集合预报控制预报,建立此次试验所用各站点 00 时 $T_{2\mathrm{m}}$ 的亚高斯 BPF 概率预报公式(预报时效为 96 h,下同)

$$P(T_k \mid X_k) = Q\{\frac{1}{t\sigma_k}[T_k - \frac{c_1}{\sigma_x}X_k - (1 - \frac{c_1}{\sigma_x})m_k - (c_0 - \frac{c_1\mu_x}{\sigma_x})\sigma_k]\} \quad (14)$$

可以看到,在此次亚高斯 BPF 初步试验中,在获得预报值 X_k 之前,对预报量 T_k 的认识仅限于先验概率分布,即气候分布 $Q(\dfrac{T_k - m_k}{\sigma_k})$;获得确定性预报值 X_k 之后,对 T_k 的认识变为正态分布形式的概率预报式(14)。由于后验参数 t 小于 1,故后验概率预报的标准差总小于气候分布,预报值 X_k 对后验概率预报的均值起调节作用。

例如,此次试验中,基于 NCEP 集合预报中的控制预报,长沙站 00 时 $T_{2\mathrm{m}}$ 的亚高斯 BPF 概率预报公式为

$$P(T_k \mid X_k) = Q[\frac{1}{0.49\sigma_k} \cdot (T_k - 0.59X_k - 0.41m_k + 0.52\sigma_k)] \quad (15)$$

基于 NMC 集合预报中的控制预报,武汉站 00 时 $T_{2\mathrm{m}}$ 的亚高斯 BPF 概率预报公式为

$$P(T_k \mid X_k) = Q[\frac{1}{0.70\sigma_k} \cdot (T_k - 0.57X_k - 0.43m_k + 0.47\sigma_k)] \quad (16)$$

4.3 独立试验

使用预报变量 2008 年 1 月 5—31 日的试验样本,建立针对不同集合预报控制预报的各站点亚高斯 BPF 概率预报公式后,利用 2008 年 2 月 1—4 日的控制预报样本进行亚高斯 BPF 概率预报试验,得出长沙站和武汉站正态分布的概率预报参数(表 1)。

表 1 长沙站和武汉站亚高斯 BPF 概率预报试验结果(℃)

Table 1 The results of the independent probabilistic forecasting experiment using the meta-Gaussian BPF at Changsha and Wuhan stations

站点	2008 年 2 月	m_k	σ_k	T_k	控制预报值			概率预报参数均值			概率预报标准差		
					NMC	ECMWF	NCEP	NMC	ECMWF	NCEP	NMC	ECMWF	NCEP
长沙站	1 日	3.1	3.2	0.3	2.2	0.2	−1.0	1.2	0.8	−1.0	2.6	2.0	1.6
	2 日	3.2	2.6	−1.8	0.2	−1.6	−2.1	0.6	−0.1	−1.3	2.1	1.6	1.3
	3 日	3.3	2.4	−3.4	−0.7	−10.9	−8.7	0.3	−5.2	−5.0	1.9	1.5	1.2
	4 日	3.1	3.2	−1.6	−1.2	−2.8	−0.2	−0.4	−0.9	−0.5	2.6	2.0	1.6
武汉站	1 日	1.3	2.4	−0.8	0.3	−1.1	−1.2	−0.4	−0.2	−1.1	1.7	1.7	0.9
	2 日	1.8	2.4	−4.7	0.1	−5.1	−2.5	−0.3	−2.1	−1.8	1.7	1.7	0.9
	3 日	2.5	2.5	−4.1	−1.0	−6.0	−1.6	−0.7	−2.2	−1.0	1.8	1.8	1.0
	4 日	2.0	2.8	−2.1	−1.5	−7.7	−0.25	−1.3	−3.4	−0.5	2.0	2.0	1.1

由表 1 可见,2008 年 2 月 1—4 日的 4 天中,长沙站 00 时 $T_{2\mathrm{m}}$ 明显低于历史同期均值。其中,2 月 3 日观测值为 −3.4℃,ECMWF、NCEP 集合预报控制预报值分别为 −10.9、−8.7℃,与之对应的亚高斯 BPF 概率预报均值分别为 −5.2 与 −5.0℃,可见,亚高斯 BPF 对控制预报极端预报值起到了明显的调节作用。对于概率预报标准差,源自 NCEP 控制预报的最小,ECMWF 的次之,NMC 的最大,这一点也可以从图 5 看出。从表中还可以看到,正如概率预报公式所表达的那样,源自 3 个控制预报的概率预报标准差都小于先验气候标准差。

从长沙站 2008 年 2 月 1 日 00 时 $T_{2\mathrm{m}}$ 的各控制预报值 X_{32} 及对应的亚高斯 BPF 正态概率密度函数形式的概率预报(图 5)可以看出,3 个概率预报所指

示的 T_{32} 比较可能出现的范围各不相同,其中源自 NCEP 控制预报的概率预报所给预报概率分布最为集中,而源自 ECMWF 控制预报的概率预报的概率密度峰值位置与观测值最为接近;还可看到,T_{32} 气候概率密度的分布比源自各控制预报的概率预报都要更分散一些。

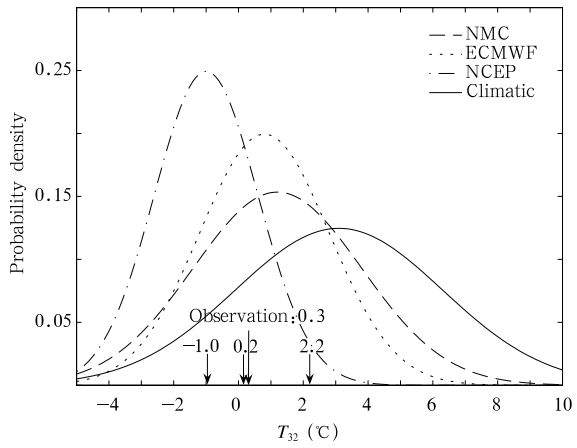


图 5 各控制预报值 X_{32} 及对应的 T_{32} 正态概率密度函数形式的概率预报(2008,长沙站)

Fig. 5 Control forecasts X_{32} from the 3 ensemble prediction centers and the corresponding probabilistic forecasts of T_{32} in the form of the normal probability density function at Changsha station in 2008

图 6 是长沙站 2008 年 2 月 1 日 00 时 T_{2m} 的 NMC 控制预报值 X_{32} 及对应的概率分布形式的亚高斯 BPF 概率预报。由连续函数形式的概率预报可以提取出多种形式的简化概率预报:(1)累积概率,由图 6,长沙站 2008 年 2 月 1 日 00 时 T_{2m} 低于 0.7℃ 的可能性为 42%;(2)定概率中央可信区间(定概率 CCI),长沙站 2008 年 2 月 1 日 00 时 T_{2m} 会以 50% 的概率出现于[−0.6℃, 3.0℃],会以 80% 的概率出现于[−2.1℃, 4.5℃],会以 95% 的概率出现于[−3.1℃, 5.5℃];(3)定宽度中央可信区间(定宽度 CCI),长沙站 2008 年 2 月 1 日 00 时 T_{2m} 出现于以概率密度峰值位置为中心的温度区间(1.2±1)℃ 中的概率为 30%,出现于(1.2±2)℃ 中的概率为 56%,出现于(1.2±3)℃ 中的概率为 75%;(4)极端概率低(高)温区间,长沙站 2008 年 2 月 1 日 00 时 T_{2m} 的 10% 极端概率低温区间为[−∞, −2.1℃],10% 极端概率高温区间为[4.5℃, ∞],5% 极端概率低温区间为[−∞, −3.1℃],5% 极端概率高温区间为[5.5℃, ∞],2.5% 极端概率低温区间为[−∞, −3.9℃],2.5% 极端概率高温区间为[6.3℃, ∞];(5)警戒低(高温)概率,设某一概率预报用户(如某一工厂或农场等)的警戒低温为 −5℃,警戒高温 10℃,根据图 6 所给概率预报,该用户的警戒低温概率为 0.9%,警戒高温概率为 0.04% ……。由此可以看到,连续函数形式的概率预报可以提供相当多种简化概率预报信息。

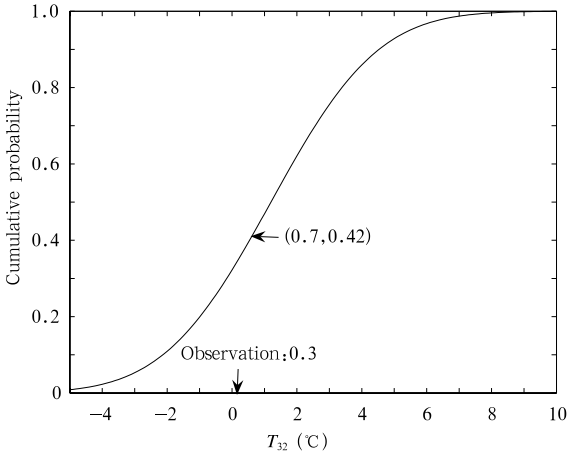


图 6 与 NMC 控制预报对应的 T_{32} 概率分布形式的概率预报(2008,长沙站)

Fig. 6 Meta-Gaussian BPF probabilistic forecast of T_{32} in the form of the cumulative probability distribution corresponding to the NMC control forecast at Changsha station in 2008

4.4 预报检验

要检验亚高斯 BPF 概率预报的性能,仅有 4 d 长度的独立预报试验结果对于概率预报检验来说显然太短。为此,利用预报因子 X_k 的所有预报样本及已建立的各亚高斯 BPF 概率预报公式,对长沙和武汉两站 2008 年 1 月 5 日—2 月 4 日(下称检验期)每日 00 时的 T_{2m} 进行亚高斯 BPF 概率预报试验,然后利用实况资料和 31 d 长度的试验结果进行检验。

针对不同类型概率预报的不同方面,现在已有很多种相应的检验方法与指标(Wilks, 2006)。根据这里的实际检验需要,选用连续分级概率评分(CRPS, Matheson, et al, 1976; Unger, 1985; Hersbach, 2000; Wilks, 2006)和亚高斯 BPF 中的有效信息评分。其中,连续分级概率评分反映的是概率预报与实况之间的差异,有效信息评分反映的是概率预报中预报因子预报能力的高低。

表 2 给出的是长沙、武汉两站各控制预报的平均绝对误差,对应的亚高斯 BPF 概率预报的均值预报平均绝对误差,以及标准差、连续分级概率评分和有效信息评分的检验期均值。从表 2 可以看出,源自各控制预报的概率预报均值预报平均绝对误差比对应的控制预报平均绝对误差要小,其中改善幅度最大的是 NMC,为1.1℃。此外,源自 3 个控制预报的概率预报的均值预报平均绝对误差和标准差均值的大小次序一致;连续分级概率评分、有效信息评分的大小次序相反,表明这两个概率预报检验指标在反映亚高斯 BPF 概率预报的性能上具有一致性。由评分的大小次序可知,在本次检验试验中,3 个概率预报在预报性能上存在差异,预报性能最高的概率预报源自 NCEP 控制预报。

表 2 长沙站和武汉站亚高斯 BPF 概率预报检验结果

Table 2 The verification results of the meta-Gaussian BPF probabilistic forecasts at Changsha and Wuhan stations

站点	预报中心	控制预报平均绝对误差(℃)	概率预报			
			均值预报的平均绝对误差(℃)	标准差(℃)	分级概率评分均值	有效信息评分均值
长沙	NWC	3.8	2.7	2.5	1.75	0.60
	ECMWF	2.3	1.8	2.0	1.23	0.78
	NCEP	2.2	1.3	1.6	0.94	0.87
武汉	NWC	3.1	2.2	2.2	1.44	0.71
	ECMWF	2.3	2.2	2.1	1.39	0.71
	NCEP	1.8	1.4	1.2	0.95	0.92

图 7 展示的检验试验结果为长沙站各控制预报亚高斯 BPF 概率预报 0.1 概率区间内的实况出现频率分布曲线。

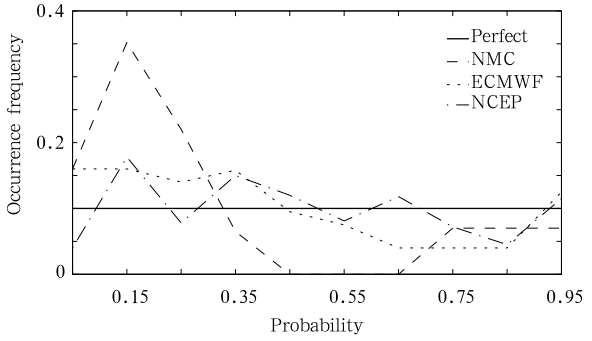


图 7 各亚高斯 BPF 概率预报按 0.1 概率间隔的实况出现频率分布

(2008 年 1 月 5 日—2 月 4 日,长沙站)

Fig. 7 Distribution of the observational occurrence frequency in the 10% probability interval of the meta-Gaussian BPF probabilistic forecast at Changsha station during 5 Jan to 4 Feb 2008

实况出现频率分布曲线是由这样一些点的连线得到的:点的纵坐标表示,在以点的横坐标为中心的概率预报 P-概率区间内的实况出现频率,这里 P 取为 0.1;它可以反映一个概率预报系统的系统偏差状况。由图 7 中与 NMC、ECMWF 控制预报对应的各概率预报实况出现频率分布曲线相对于理想频率分布直线的位置可知,由 NMC、ECMWF 的控制预报分别与对应的亚高斯 BPF 构成的两个概率预报系统都具有正(偏大)系统偏差,其中前者的更大一些;由 NCEP 控制预报与对应的亚高斯 BPF 构成的概率预报系统的系统偏差相对较小,这在表 2 中各自的连续分级概率评分的大小上都有所反映。

5 总结与讨论

本文详细介绍了一种针对确定性预报产品的概率预报统计技术——采用单一预报因子的连续型预报量亚高斯 BPF,以及对其进行的初步试验。

亚高斯似然模型的采用,使得亚高斯 BPF 较采用正态-线性似然模型的正态-线性 BPF 具有更广的适用范围。亚高斯 BPF 容许预报变量的先验概率分布、边缘概率分布是参数化的或非参数化的、正态的或非正态的;其处理的预报变量间的随机依赖结构可以具有非线性(条件均值)和异方差性(条件方差);其概率预报有 3 种连续函数形式:后验概率分布函数、后验概率密度函数和后验概率分位数函数。

利用 3 个集合预报中心控制预报资料对亚高斯 BPF 进行的初步试验结果显示,亚高斯 BPF 可以把一个确定性的 T_{2m} 预报转变为一个与之对应的 T_{2m} 概率预报;该概率预报是对那个确定性预报的预报不确定性的定量化表达。此外,初步试验结果还表明,源自不同中心控制预报的亚高斯 BPF 概率预报

的性能存在差异。

这次初步试验所采用气候、历史资料很短（T_k，20 年；X_k，1 个月）；另外，对预报变量的先验概率分布 $G(T_k)$、边缘概率分布估计 $\overline{K}(X)$所采取的正态分布近似，可能与它们的边缘概率分布之间存在着较大差距，同时正态分布近似使亚高斯 BPF 容许非正态边缘概率分布的优点没有展示出来。由此推断，此次亚高斯 BPF 概率预报试验还存在很大改进空间。

实际上，亚高斯 BPF 具有 3 种应用形式，可以分别用于连续型、二分类和多分类预报量的概率预报。此外，亚高斯 BPF 可以采用多个预报因子；如对于降水量，可选取数值预报模式的降水量预报、相对湿度预报、相对涡度预报等作为预报因子，利用由它们构成的一个预报因子向量进行定量降水概率预报（Krzysztofowicz, et al, 2006b）。由于具有应用上的很多优点，亚高斯 BPF 正逐渐得到重视和推广应用。Krzysztofowicz (2007)新的研究目标——针对集合预报的贝叶斯集合预报处理器（BPE）的技术基础正是亚高斯 BPF。

致谢：感谢 NCEP 朱跃建在文献搜索方面提供的帮助以及四川省气象台陈朝平提供的意见和建议，同时感谢国家气象中心预报系统开放实验室刘小魏提供观测资料。

参考文献

陈超辉，李崇银，谭言科等. 2010. 基于交叉验证的多模式超级集合预报方法研究. 气象学报，68(4)：464-476

陈朝平，冯汉中，陈静. 2010. 基于贝叶斯方法的四川暴雨集合概率预报产品释用. 气象，36(5)：32-39

陈法敬，矫梅燕，陈静. 2011. 一种温度集合预报产品释用方法的初步研究. 气象，37(1)：14-20

陈静，陈德辉，颜宏. 2002. 集合数值预报发展与研究进展. 应用气象学报，13(4)：497-507

范新岗. 1999. 集合预报方法的全局研究. 气象学报，57(1)：74-83

黄嘉佑. 2000. 气象统计分析与预报方法（第 2 版）. 北京：气象出版社，246pp

矫梅燕，龚建东，周兵等. 2006. 天气预报的业务技术发展. 应用气象学报，17(5)：594-601

李泽椿，陈德辉. 2002. 国家气象中心集合数值预报业务系统的发展及应用. 应用气象学报，13(1)：1-15

刘金达. 2000. 集合预报开创了业务数值天气预报的新纪元. 气象，26(6)：21-25

马培迎. 1999. 应用贝叶斯原理修正降水概率预报. 气象科技，1：45-48

盛骤，谢式千，潘承毅. 2001. 概率论与数理统计（第 3 版）. 北京：高等教育出版社，434pp

王善序. 2001. 贝叶斯概率水文预报简介. 水文，21(5)：33-34

薛纪善. 2006. 新世纪初我国数值天气预报的科技创新研究. 应用气象学报，17(5)：602-610

幺枕生，丁裕国. 1990. 气候统计（修订本）. 北京：气象出版社，102pp

俞小鼎，张艺萍. 2004. 天气预报技巧和价值的关系. 气象科技，32(6)：393-398

张大林. 2005. 大气科学的世纪进展与未来展望. 气象学报，63(5)：812-824

张洪刚，郭生练，刘攀等. 2004. 基于贝叶斯分析的概率洪水预报模型研究. 水电能源科学，22(1)：22-25

张洪刚，郭生练，何新林等. 2006. 水文预报不确定性的研究进展与展望. 石河子大学学报（自然科学版），4(1)：15-21

张军，葛军，田俊杰等. 1998. 概率天气预报及其应用. 北京：气象出版社，168pp

周秀骥. 2005. 大气随机动力学与可预报性. 气象学报，63(5)：806-811

Abramowitz M, Stegun I A. 1972. Handbook of mathematical functions. Dover, Mineola, New York, 931pp

Antolik M S. 2000. An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. J Hydrol, 239(1－4)：306-337

Glahn H R, Lowry D A. 1972. The use of model output statistics (MOS) in objective weather forecasting. J Appl Meteor, 11(8)：1203-1211

Herr H D, Krzysztofowicz R. 2005. Generic probability distribution of rainfall in space: the bivariate model. J Hydrol, 306：234-263

Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Wea Forecasting, 15：559-570

Kelly K S, Krzysztofowicz R. 1995. Bayesian Revision of an arbitrary prior density. Proceedings of the Section on Bayesian Statistical Science, American Statistical Association, 50-53

Kelly K S, Krzysztofowicz R. 1997. A bivariate meta-Gaussian density for use in hydrology. Stochastic Hydrology Hydraulics, 11: 17-31

Krzysztofowicz R. 1983. Why should a forecaster and decision maker use Bayes theorem. Water Resour Res, 19：327-336

Krzysztofowicz R. 1987. Markovian forecast processes. J American Statistical Association, 82：31-37

Krzysztofowicz R. 1992. Bayesian correlation score: A utilitarian measure of forecast skill. Mon Wea Rev, 120(1)：208-219

Krzysztofowicz R. 1997. Transformation and normalization of variates with specified distributions. J Hydrol, 197：286-292

Krzysztofowicz R. 2002. New statistical techniques for probabilistic weather forecasting—stage Ⅰ: Project summary. Department of System Engineering, University of Virginia, 1pp

Krzysztofowicz R. 2007. New statistical techniques for probabilistic weather forecasting—stage Ⅱ: Project summary. Department of System Engineering, University of Virginia, 1pp

Krzysztofowicz R. 2008. Probabilistic forecasts from the national digital forecast database. Wea Forecasting, 23:270-289

Krzysztofowicz R, Kelly K S. 2000a. Bayesian improver of a distribution. Stochastic Environmental Res Risk Assessment, 14(6):449-470

Krzysztofowicz R, Kelly K S. 2000b. Hydrologic uncertainty processor for probabilistic river stage forecasting. Water Resour Res, 36:3265-3277

Krzysztofowicz R, Maranzano C J. 2006a. Bayesian processor of output:Probability of precipitation occurrence. http://faculty.virginia.edu/rk/BPO.htm

Krzysztofowicz R, Maranzano C J. 2006b. Bayesian processor of output: Probabilistic quantitative precipitation forecast. http://faculty.virginia.edu/rk/BPO.htm

Krzysztofowicz R, Reese S. 1991. Bayesian analysis of seasonal runoff forecasts. Stochastic Hydrology Hydraulics, 5(4):267-279

Krzysztofowicz R, Watada L M. 1986. Stochastic model of seasonal runoff forecasts. Water Resour Res, 22:296-302

Lehmann E L. 1966. Some concepts of dependence. Annals Mathematical Statistics, 37:1137-1153

Matheson J E, Winkler R L. 1976. Scoring rules for continuous probability distributions. Management Sci, 22:1087-1096

Unger D A. 1985. A method to estimate the continuous ranked probability score// Preprints, 9th Conference on Probability and Statistics in the atmospheric Sciences. Amer Meteor Soci, 206-213

Wilks D S. 2006. Statistical Methods in the Atmospheric Sciences (Second edition). UK:Elsevier, 302pp

附 录

正态分布 Q 及其逆 Q^{-1} 的近似表达式

$$Q(x)\approx\eta(x)=\begin{cases}1-\dfrac{a_1+a_2t^2+a_3t^3}{\sqrt{2\pi}}\exp(-\dfrac{x^2}{2}) & x>0\\ \dfrac{a_1+a_2t^2+a_3t^3}{\sqrt{2\pi}}\exp(-\dfrac{x^2}{2}) & x\leqslant 0\end{cases}$$

其中，$t=\dfrac{1}{1+a\mid x\mid}$，$a=0.33267$，$a_1=0.436183$，$a_2=-0.1201676$，$a_3=0.9372986$。近似表达式 $\eta(x)$ 的绝对误差，$\mid Q(x)-\eta(x)\mid<10^{-5}$（盛骤等，2001）。

$$Q^{-1}(p)\approx z(p)=\begin{cases}\dfrac{a_0+a_1t}{1+b_1t+b_2t^2}-t & t=\sqrt{-2\ln p} \quad 0<p<0.5\\ t-\dfrac{a_0+a_1t}{1+b_1t+b_2t^2} & t=\sqrt{-2\ln(1-p)} \quad 0.5\leqslant p<1\end{cases}$$

其中，$a_0=2.30753$，$a_1=-0.27061$，$b_1=0.99229$，$b_2=0.04481$。近似表达式 $z(p)$ 的绝对误差，$|Q^{-1}(p)-z(p)|<3\times10^{-3}$（Abramowitz, et al, 1972; Krzysztofowicz, et al, 2006a）。