

华连生, 温华洋, 朱华亮, 等. 2015. 基于 Bayes 判别法的霜生自动化观测模型探讨. 气象, 41(8): 964-969.

基于 Bayes 判别法的霜生自动化 观测模型探讨^{*1}

华连生¹ 温华洋¹ 朱华亮¹ 张正铨²

1 安徽省气象信息中心, 合肥 230031

2 中国科学技术大学数学学院, 合肥 230022

提 要: 利用安徽砀山气象站的 2001—2013 年冬半年(10 月至次年 4 月)的观测资料, 探讨霜生与气温、地温、水汽压和风速等气象要素的相关性, 并基于 Bayes 判别方法, 采用逐步判别分析, 建立多套霜生自动判别模型。结果表明:(1)霜是否出现与日最低及夜间不同观测时次的气温、地表温度显著相关, 当夜间气温或地表温度越低, 低于霜点的可能性越大, 结霜的可能性也越大。(2)通过回算性检验和独立样本的预报性检验, 基于 Bayes 判别法的霜生模型, 对霜未发生的平均判别准确率达到 86.5%, 对霜发生的平均判别准确率达到 92.7%, 其中用日最低地温、当日 07 时水汽压和当日 07 时风速所建立的三要素模型最优, 对霜发生的判别准确率可达到 90%以上。因此, 可以将 Bayes 霜生判别模型与图像识别技术相结合应用于霜的自动化观测。

关键词: 霜生, Bayes 判别分析, 自动化观测

中图分类号: P412.423

文献标志码: A

doi: 10.7519/j.issn.1000-0526.2015.08.005

Automated Observation Model for Frost Based on Bayes Discriminant Method

HUA Liansheng¹ WEN Huayang¹ ZHU Hualiang¹ ZHANG Zhengquan²

1 Anhui Meteorological Information Centre, Hefei 230031

2 School of Mathematics, University of Science and Technology of China, Hefei 230022

Abstract: The correlations between frost and temperature, surface temperature, vapor pressure, wind speed and other meteorological factors are discussed in this paper by using the observation data from Anhui Dangshan Weather Station in the winter half-year (from October to April of the next year) from 2001 to 2013. Using stepwise discriminant analysis method, multiple sets of frost automatic discriminant models for the occurrence of frost are built based on Bayes discriminant method. The results show that: (1) The occurrence of frost is significantly correlated with daily minimum temperature, night temperature of different observation time and surface temperature. The lower the night temperature or the surface temperature is, the larger the possibility of the temperature is lower than the frost point and the greater the possibility of the frost occurrence. (2) Through the back calculation test and prediction test of independent samples, the average accuracy rate of un-occurred frost discriminated by the frost model is 86.5% based on Bayes discriminant method and the average accuracy rate of the seen frost is 92.7%. The three factor models based on the daily minimum temperature, the daily vapor pressure at 07:00 BT and the daily wind speed at 07:00 BT are optimal. The accuracy rate of discriminating the frost occurrence by the three factor models

* 安徽省气象科技发展基金(KM201405)资助

2014 年 6 月 9 日收稿; 2015 年 2 月 27 日收修定稿

第一作者: 华连生, 主要从事气象数据处理和信息网络业务应用系统设计. Email: ahhls@sina.com

is more than 90%. Therefore, we can combine the Bayes frost discriminant model with image recognition technology, and apply the new technology to frost automatic observation.

Key words: frost occurrence, Bayes discriminant analysis, automatic observation

引 言

根据地面气象观测规范(中国气象局,2003)的定义,霜是指水汽在地面和近地面物体上凝华而成的白色松脆的冰晶。易在晴朗风小且寒冷的夜间生成,亦即霜的生成需要适度的风速、温度、湿度条件:即气温降到霜点时,若地面或近地面物体降到 0°C 以下,水汽过饱和部分在地面凝华或冻结,形成霜,但若温度过高,湿度过大将会先形成大量的露,而露的凝结过程必将伴随着凝结潜热的释放,将部分补偿了因辐射而损失的热量,不利于霜的生成,另外风可以增加水汽输送率,使凝结加快,但若风速过大,影响水汽的聚集凝华,同样不利于霜的生成。(盛裴轩等,2003;王国复等,2009;李辑等,2010)。

秋季早霜和春季晚霜都会对植物(特别是农作物)产生冻害。春霜会对小麦产生危害(许莹等,2014),晚霜会对果树的嫩芽、幼叶、花和幼果造成危害,严重时会造成果物大幅度减产,甚至绝收,经济损失巨大(金波,2013;张雪芬等,2012),属于主要低温冷冻害之一。如1995年9月华北地区北部及黑龙江、吉林、陕西、宁夏等省(区)的部分地区遭遇初霜冻害,有248万 hm^2 农作物受灾,直接经济损失近10亿元(Wang et al,2008)。因此,气象部门对霜的监测和霜冻的预报预测非常重视。地面气象观测规范(中国气象局,2003)要求对霜随时进行记录,对霜冻灾害进行调查记载。此外,在气象业务也已经形成一套较成熟的霜冻预报预测技术方法(林日暖等,2003;朱其文等,2003;温显罡等,2012),来满足日常气象服务和气象灾害服务的需求。

21世纪以来,随着科学技术的发展,我国气象观测自动化迅猛推进,已从气温、气压、湿度、降水等基本气象要素拓展到云、能见度和天气现象。自动气象观测能够节约大量人力、物力和财力,同时能够避免人工观测存在的主观性强、观测频率低、恶劣环境无法观测等诸多弊端。针对霜的自动化观测,国内目前主要基于示踪物体的图像、光学散射等物理

特征开展自动识别(陈冬冬等,2011;马舒庆等,2011)。由于相关设备价格较高,且自动识别的准确性还不理想,如北京气象台2010年8—10月采用DZT-1型天气现象自动化观测仪对凝结天气现象进行了自动观测试验,准确率分别仅有84%、83.3%和71%(马舒庆等,2011),不利于大量装备和全面业务化。

霜是在一定的天气条件下产生的,是各气象要素变化的综合结果。提高霜的自动化识别的准确性,还得从霜生的天气条件入手。目前风速、温度、湿度等气象要素都已实现了自动化观测,那么能否以此为基础,借鉴霜冻预报预测技术方法来探究霜的自动化观测呢?本文旨在利用安徽砀山站2001—2013年有霜期的气象观测资料,基于霜生与气象要素的相关性,运用Bayes判别法来探讨相应的自动观测模型,为全面开展霜的自动化观测提供参考。

1 资 料

砀山站位于安徽省北部,1954年10月建站,对气温等基本气象要素进行每天4次观测,2001年启动观测自动化后,对气温等基本气象要素进行每天24次观测。因此,为探讨霜生与夜间各个观测时次气象要素的相关性,本文使用的资料时段为2001—2013年,其中2001—2011年的逐日观测资料主要用于建模,2011—2013年的观测资料用于效果检验。

表1为2001—2013年砀山站的初霜日、终霜日和有霜期实况,其中平均有霜期为164 d,最早初霜为10月18日,最晚终霜为4月27日。因此,为增强研究的针对性,霜生建模时实际使用的资料为历年10月1日至次年4月30日。霜及相关气象要素分辨率见表2,其中霜为定性观测,0表示当日不出现霜,1表示当日出现霜。各要素实测资料均通过质量控制,资料完整性接近100%。

表1 2001—2013年砾山站初霜、终霜和有霜期实况

Table 1 The early frost, late frost and frost period at Dangshan Station over 2001—2013

年份	初霜日	终霜日	有霜期/d	年份	初霜日	终霜日	有霜期/d
2001—2002	10月30日	4月9日	162	2002—2003	10月26日	3月22日	148
2003—2004	10月23日	4月4日	165	2004—2005	10月22日	3月31日	161
2005—2006	10月23日	4月20日	180	2006—2007	11月6日	4月4日	150
2007—2008	10月30日	4月3日	157	2008—2009	11月9日	4月2日	145
2009—2010	11月2日	4月27日	177	2010—2011	10月27日	4月19日	175
2011—2012	10月25日	4月3日	162	2012—2013	10月18日	4月21日	186
平均状况	10月28日	4月12日	164				

表2 霜及相关气象要素分辨率

Table 2 Frost and related meteorological factor resolution

气象要素	记号	观测分辨率
气温/°C	T	日最低(min)、日逐小时(02—08时)
地表温度/°C	D	日最低(min)、日逐小时(02—08时)
水汽压/hPa	e	日逐小时(02—08时)
风速/m·s ⁻¹	F	日逐小时(02—08时)
霜	Y	日霜是否出现(0表示不出现, 1表示出现)

2 霜与气象要素的相关性分析

霜是各气象要素变化的综合结果,但是不同气象要素对霜生的影响程度应该有所不同,差异可以用相关系数来简单辨别。为此,本文利用上述观测

数据,计算了日霜是否出现与当日不同观测时次的气温(日最低气温、02—08逐小时气温)、地表温度(日最低地温、02—08时逐小时地温)、水汽压(02—08时逐小时水汽压)、风速(02—08时逐小时风速)的相关系数。如表3所示,日霜是否出现与日最低及不同观测时次的气温、地表温度有非常显著的相关性,相关系数在-0.748~-0.712,通过 $\alpha=0.01$ 的显著性水平检验,这表明温度对霜的形成影响较大,当夜间气温或地表温度越低,低于霜点的可能性越大,该日结霜的可能性自然就大。就温度要素与霜的相关系数而言,夜间不同时段差异并不显著,可以选择日最低气温和日最低地表温度作为代表,来研究其对霜生的影响。

表3中水汽压与日霜是否出现呈现较显著的负

表3 霜与各气象要素的相关系数

Table 3 The correlation coefficient of frost and meteorological elements

	02时	03时	04时	05时	06时	07时	08时	日最低
$R_{(T,Y)}$	-0.717	-0.721	-0.722	-0.726	-0.728	-0.733	-0.712	-0.730
$R_{(D,Y)}$	-0.741	-0.742	-0.743	-0.744	-0.744	-0.748	-0.729	-0.731
$R_{(e,Y)}$	-0.634	-0.636	-0.633	-0.633	-0.632	-0.632	-0.613	—
$R_{(F,Y)}$	-0.166	-0.156	-0.144	-0.144	-0.157	-0.172	-0.220	—

注:Y、T、D、e和F代表的要素同表2;各相关系数均通过 $\alpha=0.01$ 的显著性水平检验。

相关关系,表明并不是水汽越充足越有利于霜的形成,而是要看温度是否在霜点以下,且局部水汽是否达到饱和(即相对湿度达到100%)而析出。由于饱和水汽压与温度正相关,并随温度的降低而迅速减小,因此,水汽压偏低既可能是水汽少,也可能是温度低。若是后者,表明结霜的几率会增大,所以简单统计得到的水汽压与日霜是否出现的相关关系可能带有较强的温度变化信号。此外,风对于霜的形成也有影响。有微风的时候,空气缓慢地流过冷物体表面,不断地供应着水汽,有利于霜的形成。相反,风大的时候,由于空气流动得很快,接触冷物体表面的时间太短,同时风速大的时候,上下层的空气容易

互相混合,不利于温度降低,从而也会妨碍霜的形成,所以风速与日霜是否出现的相关系数也为负值。

3 基于Bayes判别法的霜生模型

3.1 Bayes判别法

Bayes判别分析是多变量统计分析中用于判别样品所属类型的一种统计分析方法。其在判别过程中既要给出一个衡量新样品与各已知类别接近程度的统计模型,也需指定一种判别规则(王斌会,2011;高慧璇,2005)。本文中,它所要解决的问题是根据

已有气象资料(温度、水汽压等)来判断某日是有霜还是无霜。

Bayes 判别属于经典的概率性判别分析方法,通常要知道先验概率,即事先知道日霜是否出现的概率,然后根据某一天的气象观测资料计算该日无霜和有霜的条件概率 $P(j/x)$, $j=1,2$ 。其中 $P(1/x)$ 表示该日无霜的条件概率, $P(2/x)$ 表示该日有霜的条件概率, x 表示该日的相关要素(温度、水汽压等)的观测值。比较 $P(1/x)$ 和 $P(2/x)$ 的大小,若 $P(1/x)$ 大就认为该日无霜出现,反之认为该日有霜出现。简言之,基于 Bayes 判别法的霜生模型就是要在已知日霜出现和不出现的概率及当日相关气象要素观测值条件下,分别求出当日有霜现象和无霜现象时的条件概率,然后根据概率大小判定霜是否出现。

假设日霜出现和不出现的概率分别为 q_1 和 q_2 , 无霜类和有关霜类的概率密度函数分别为 $P_1(x)$ 和 $P_2(x)$, x 表示该日的相关要素(温度、水汽压等)观测值。则该日有霜或无霜的条件概率可表示为:

$$P(j/x) = \frac{q_j p_j(x)}{q_1 p_1(x) + q_2 p_2(x)}, \quad j=1,2 \quad (1)$$

当 $P(1/x) = \max_{1 \leq j \leq 2} P(j/x)$ 时,认为当日无霜;反之,认为当日有霜。若无霜类和有关霜类总体的协方差矩阵相同,本文利用数学手段(王斌会,2011)将式(1)简化为如下显示的线性判别函数:

$$\begin{cases} Y_1 = \ln q_1 + a_{00} + a_{01} X_1 + a_{02} X_2 + \dots + a_{0p} X_p \\ Y_2 = \ln q_2 + a_{10} + a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p \end{cases} \quad (2)$$

式中, $a_{i0}, a_{i1}, \dots, a_{ip}$ ($i=1,2$) 为判别系数, Y_1 代表无霜类, Y_2 代表有霜类。把实际测得的气象要素值 X_1, \dots, X_p 代入式(2),可求得 Y_1 和 Y_2 值。若 $Y_1 > Y_2$, 认为当日无霜;反之,认为当日有霜。

3.2 基于 Bayes 判别法的霜生模型

本文用 2001—2011 年期间霜生的频率来代替先验概率 q_1 和 q_2 。建模数据中共有 2122 d 的观测值,其中无霜日 1209 d,有霜日 913 d,无霜发生的频率 $q_1=0.570$,有霜发生的频率 $q_2=0.430$ 。

鉴于上述霜与气象要素的相关性分析,本文采用逐步判别的思路将气温/地温、水汽压、风速依次代入式(2)建模,并考查霜生模型的判别能力。另据统计,10月至次年4月砭山日最低气温、日最低地温多出现在 06—07 时(图略)。结合表 3 的相关系

数值的大小,在建模过程中,最终选择 07 时水汽压、07 时风速与日最低气温、日最低地温进行计算。

第一步,单一的温度判别模型。仅将日最低气温(T_{\min})和日最低地温(D_{\min})分别代入式(2),利用 SPSS 软件(陈胜可,2010)可得到式(3)和式(4),即模型 1(日最低气温)和模型 2(日最低地温)。

模型 1:

$$\begin{cases} Y_1 = 0.372 T_{\min} - 1.866 \\ Y_2 = -0.139 T_{\min} - 1.026 \end{cases} \quad (3)$$

模型 2:

$$\begin{cases} Y_1 = 0.352 D_{\min} - 1.630 \\ Y_2 = -0.179 D_{\min} - 1.120 \end{cases} \quad (4)$$

第二步,温度和水汽判别模型。将 07 时水汽压(e_{07})和日最低气温(T_{\min})作为自变量代入式(2)建立判别模型,得到模型 3 即为式(5),同样将 07 时水汽压(e_{07})和日最低地温(D_{\min})作为自变量代入式(2)建立判别模型,得到模型 4 即为式(6)。

模型 3:

$$\begin{cases} Y_1 = -0.724 T_{\min} + 1.844 e_{07} - 6.941 \\ Y_2 = -1.327 T_{\min} + 1.997 e_{07} - 6.983 \end{cases} \quad (5)$$

模型 4:

$$\begin{cases} Y_1 = -1.084 D_{\min} + 2.243 e_{07} - 8.122 \\ Y_2 = -1.761 D_{\min} + 2.472 e_{07} - 9.004 \end{cases} \quad (6)$$

第三步,温度、水汽和风判别模型。将 07 时水汽压(e_{07})、风速(F_{07})和日最低气温(T_{\min})作为自变量代入式(2)建立判别模型,得到模型 5 即为式(7),同样将与 07 时水汽压(e_{07})、风速(F_{07})和日最低地温(D_{\min})作为自变量代入式(2)建立判别模型,得到模型 6 即为式(8)。

模型 5:

$$\begin{cases} Y_1 = -0.948 T_{\min} + 2.284 e_{07} + 2.277 F_{07} - 10.279 \\ Y_2 = -1.516 T_{\min} + 2.362 e_{07} + 1.882 F_{07} - 9.257 \end{cases} \quad (7)$$

模型 6:

$$\begin{cases} Y_1 = -1.217 T_{\min} + 2.566 e_{07} + 2.135 F_{07} - 11.150 \\ Y_2 = -1.864 T_{\min} + 2.718 e_{07} + 1.636 F_{07} - 10.773 \end{cases} \quad (8)$$

3.3 霜生模型的效果检验

为检验各模型的判别效果,本文采用双向检验,即利用 2001—2011 年上半年建模数据回算性检验和 2011 年下半年至 2013 年观测数据预报性检验,其中后者属于独立检验样本。

表 4 给出了各模型回算结果,其中建模数据中无霜日为 1209 d,有霜日为 913 d,共计 2122 d。就模型 1 而言,1209 个无霜日中判别出无霜日 1040 d,对霜未发生的判断正确率为 86.0%,913 个有霜日判别出有霜日 823 d,对霜发生的判断正确率为 90.1%,而模型 1 的整体判断正确率达到 87.8%。这表明只考虑日最低气温所建立的模型 1

是合理的,可以接受的。而对于只考虑日最低地表温度的模型 2,其整体判断正确率可以达到 88.5%,要优于模型 1,但其霜未发生判断正确率要低于模型 1。将次要因素水汽和风逐步分别加入模型 1 和 2 中,所得到模型均有较高的判断正确率,且运用日最低地表温度所建立的模型随着要素的增加判断正确率在逐步提高。

表 4 基于 2001—2011 年观测数据的各模型回算结果

Table 4 The back calculation results for each model based on observation data over 2001—2011

台站	霜未发生 日数/d	霜未发生判断 正确日数/d	未发生判断 正确率/%	霜发生 日数/d	霜发生判断 正确日数/d	发生判断 正确率/%	整体判断 正确率/%
模型 1	1209	1040	86.0	913	823	90.1	87.8
模型 2	1209	1003	83.0	913	876	95.9	88.5
模型 3	1209	1048	86.7	913	811	88.8	87.6
模型 4	1209	1017	84.1	913	868	95.1	88.8
模型 5	1209	1059	87.6	913	815	89.3	88.3
模型 6	1209	1051	86.9	913	861	94.3	90.1

表 5 给出了 2011—2013 年观测数据的各模型检验结果,其中无霜日为 232 d,有霜日为 193 d,共计 425 d。模型 1 对 232 个无霜日中判别出无霜日 204 d,对霜未发生的判断正确率为 87.9%,193 个有霜日中判别出有霜日 172 d,对霜发生的判断正确率为 89.1%,而整体的判断正确率达到 88.5%。从检验效果来看,其他模型同样对独立样本具有较高判断正确率,且运用日最低地表温度所建立的模型要优于日最低气温所建立的模型。经过回算性检验和预报性检验发现,所建立的 6 个霜生判别模型对霜未发生的平均判别准确率达到 86.5%,霜发生的

平均判别准确率达到 92.7%,且气温类模型(模型 1、3 和 5)霜未发生的判断正确率略高于对应的地面温度类模型(模型 2、4 和 6),霜发生判断的准确率则反之。在上述模型中由用日最低地温、当日 07 时水汽压和当日 07 时风速所建立的三要素模型最优,对霜发生的判别准确率可达到 90%以上,但我们发现不论是有霜日还是无霜日,所建模型都有判错的可能性。为此,还要探究其他方法与之结合,来提高天气现象霜的预测准确率,使之更加接近 100%,从而应用于气象自动观测。

表 5 基于 2011—2013 年观测数据的各模型检验结果

Table 5 The inspection results for each model based on observation data over 2011—2013

台站	霜未发生 日数/d	霜未发生判断 正确日数/d	未发生判断 正确率/%	霜发生 日数/d	霜发生判断 正确日数/d	发生判断 正确率/%	整体判断 正确率/%
模型 1	232	204	87.9	193	172	89.1	88.5
模型 2	232	192	82.8	193	189	97.9	89.7
模型 3	232	207	89.2	193	168	87.1	88.2
模型 4	232	197	84.9	193	187	96.9	90.4
模型 5	232	211	91.0	193	177	91.7	91.3
模型 6	232	204	87.9	193	186	96.4	91.8

4 结论与讨论

(1) 霜是在一定的天气条件下产生的,其是否出现与日最低及不同观测时次的气温、地表温度有非常显著的相关性,相关系数在 $-0.748 \sim -0.712$,

表明夜间气温或地表温度越低,低于霜点的可能性越大,结霜的可能性也越大。就温度要素与霜的相关系数而言,夜间不同时次间差异并不显著。

(2) 基于 Bayes 判别法和逐步判别的思路,分别建立 6 个霜生判别模型,利用建模数据的回算性检验和独立样本的预报性检验,表明霜未发生的平

均判别准确率达到 86.5%,霜发生的平均判别准确率达到 92.7%,其中由用日最低地温、当日 07 时水汽压和当日 07 时风速所建立的三要素模型最优,对霜发生的判别准确率可达到 90%以上。气温类模型和地温类模型对霜发生和未发生判断的敏感性存在一定差异。

本文是借鉴霜冻预报预测技术方法来探究霜的自动化观测,判别准确性较北京气象台 2010 年采用 DZT-1 型天气现象自动化观测仪略高,但是为提高霜的判别准确率,还可以将 Bayes 判别模型与图像识别技术相结合,应用于自动气象观测。当两者都判定有霜发生时,就认定当日有霜出现;当两者都判断无霜发生时,就认定当日无霜出现;若两者判断结果相反时,进行人工判定。

致谢:本文得到国家信息中心周自江老师的悉心指导,特此致谢!

参考文献

- 陈冬冬,施丽娟,李肖霞,等. 2011. 天气现象自动化观测现状调研. 气象科技, 39(5):596-602.
- 陈胜可. 2010. SPSS 统计分析从入门到精通. 北京:清华大学出版社, 331-341.
- 高慧璇. 2005. 应用多元统计分析. 北京:北京大学出版社, 175-210.
- 金波. 2013. 北方果树晚霜危害及防治措施. 北方果树, (3):40-41.
- 李辑,严晓瑜,王颖. 2010. 辽宁省近 50 年霜的气候变化特征. 气象, 36(11):38-45.
- 林日暖,崔巧娟,朱正心. 2003. 广东经济林果寒害地面预警强信号和长期统计预报模式的研究. 应用气象学报, 14(4):499-501.
- 马舒庆,吴可军,陈冬冬,等. 2011. 天气现象自动化观测系统设计. 气象, 37(9):1166-1172.
- 盛裴轩,毛节泰,李建国,等. 2003. 大气物理学. 北京:北京大学出版社, 310-325.
- 王斌会. 2011. 多元统计分析及 R 语言建模. 广州:暨南大学出版社, 131-153.
- 王国复,许艳,朱燕君,等. 2009. 近 50 年我国霜期的时空分布及变化趋势分析. 气象, 35(7):61-67.
- 温显昱,马舒庆,杜波,等. 2012. 草温与露点温度对结露和结霜指示性分析. 气象, (6):745-750.
- 许莹,马晓群,王晓东,等. 2014. 安徽省冬小麦春霜冻害气象指标的研究. 气象, 40(7):852-859.
- 张雪芬,余卫东,王春乙. 2012. 基于作物模型灾损识别的黄淮区域冬小麦晚霜冻风险评估. 高原气象, (1):277-284.
- 中国气象局. 2003. 地面气象观测规范. 北京:气象出版社, 23-26.
- 朱其文,张丽,孙霞. 2003. 吉林省初霜中长期预报方法研究. 吉林气象, (增刊):22-25.
- Wang Yongguang, Zhou Zijiang, Zhang Qiang, et al. 2008. Main meteorological disasters and their impacts on the economic and societal developments in China. Int J Risk Assessment Management, 8(4):384-394.