

夏季地面温度场月预报的统计模式*

陈英仪 张秋庆

(国家海洋环境预报中心, 北京 100081)

提 要

用 15 年北半球夏季地面温度的资料分析了月平均场与该月前一周左右的逐日平均场在时间和波数上的相关。发现越新、越靠后的日平均场与该月的平均场相关越强。用第 0 天外推的未来 30 天平均场与实况的相关比用前 30 天平均场作外推的相关要显著得多。由此得出结论认为各种预报方法的效果检验应与前者定义的惯性预报作比较。

文中提出了几个用于作月预报的统计模式, 分别用不同时间或不同波数作预报因子。其结果均使预报准确率有所提高。既用不同时间又用不同波数作预报因子的模式更为理想。

文中最后讨论了一个源于动力学考虑的统计模式。近 1400 个月预报例子的结果与实况的平均相关系数高达 0.75, 显著优于惯性预报。

关键词: 地面温度; 长期预报; 统计预报; 统计动力模式。

一、引 言

在短期天气预报中, 统计预报方法已逐渐被数值预报所代替。然而人们试图在长期天气预报上走同样的路时却遇到了困难。不论是理论上还是实践上, 长期预报的问题至今尚未解决。预报仍停留在相当低的水平。对一些预报水平能超过惯性预报的长期预报方法进行研究, 在目前仍是非常有意义的。

就目前情况来说, 在长期预报中统计方法比数值方法有一些优越条件。首先, 它不存在由于初始场的微小误差随积分时间的延长而急剧增长的问题。其次, 它能充分利用历史资料, 使模式能较好地反映天气形势的历史演变。因此, 统计方法至今还是长期预报的重要手段。

本文针对夏季地面温度场, 先对其月平均形势与逐月平均场的关系作分析, 建立一些统计预报模式, 比较其预报效果, 并向读者推荐一个预报效果颇优的统计—动力模式。

二、资 料

本文所用资料为 15 年 (1967—1981) 夏季 123 天 (每年从 5 月 15 日至 9 月 14 日) 每日两次的北半球 1000 hPa 温度场。资料用球谐函数的系数, 在波数 15 处作三角

1992 年 6 月 8 日收到, 1992 年 8 月 12 日收到修改稿。

* 本文是国家自然科学基金的资助项目。

截断, 并把它处理成对赤道对称而记录的。这些资料由美国 Scripps 海洋研究所 J. Roads 提供, 它来源于文献[1]。

1000 hPa 温度的物理空间分布与球谐系数之间的关系为

$$T(\lambda, \varphi, t) = \sum_{n=0}^{15} \sum_{m=0}^{15} \{T_{c_n}^m(t) \cos m\lambda + T_{s_n}^m(t) \sin m\lambda\} \cdot P_n^m(\sin \varphi), \quad (1)$$

其中 λ 、 φ 和 t 分别为经度、纬度和时间, P_n^m 为伴随勒让德函数, n 和 m 为二纬波数和纬向波数。由于对赤道对称, $(n-m)$ 为奇数时, $T_{c_n}^m$ 和 $T_{s_n}^m$ 等于零。当 $m=0$ 时, $T_{s_n}^m=0$ 。因此每个 T 的记录共有 136 个非零系数, 其中 $T_{c_n}^m$ 有 72 个, $T_{s_n}^m$ 有 64 个。这些系数与空间温度场 $T(\lambda, \varphi)$ 之间很容易相互转换。

把上述所有记录取平均作为夏季地面温度场的气候值, 再用原始记录减去此气候值得到距平值。作为预报问题, 就是要求出相应的 136 个系数并与实况比较。为简化起见, 以下的 T 都略去上、下标。应注意, 它表示的是球谐函数的系数。

三、月平均场与日平均场的关系

图 1 的实线 $\hat{T}(x)$ 表示月平均场与其前后一周左右的日平均场的相关关系。纵坐标为两者的流型相关系数 ρ , 横坐标 $x=0$ 为某月 (以 30 天计) 的前一天, x 的正值表示该月的第 x 天, 负值表示该月的前 $(x+1)$ 天。由于每年有 123 天, 可组成 93 个连续的 30 天, 所以图中的相关系数是 15×93 个例子的平均值。由图可见, 离某月的日期越远的日平均与该月平均场的相关性越差。作为预报问题, 只能用以前的资料预报以后的天气。图 1 表明以该月前一天的形势与之相关最大, 这时的 ρ 值为 0.58。若对其前后一周左右取不同时段平均, 再与月平均场求相关, 结果表示在图 1 的虚线 $\hat{T}(0-x)$ 中。这时 x 对应的 ρ 值为用第 x 天至第 0 天之平均场与相应的月平均场的相关系数。虚线与实线比较发现, 多日平均后的场比某一天的场更接近月平均场。尤其是在 30 天内取若干大的平均, 它与该月平均的相关甚高。假如允许过若干天发布该月月预报的话, 只要取一下平均, 过一天后作预报的准确率可达 0.64, 过 2 天作预报可达 0.68。如果只能用过去的资料, 则取平均的时段不能过长, 约取前 3—4 天的平均即可。若

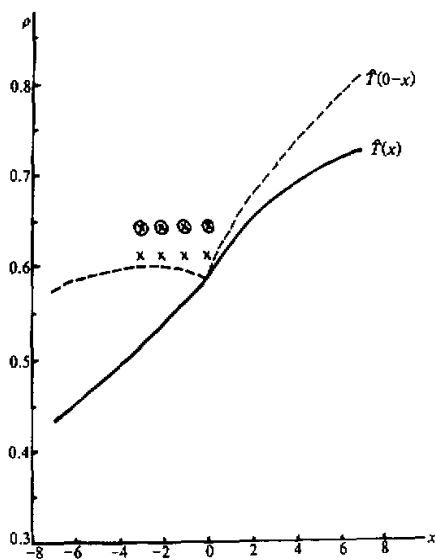


图 1 日平均场与月平均场的相关关系
实线 $\hat{T}(x)$ 为第 x 天的日平均场与该月平均场的关系;
虚线 $\hat{T}(0-x)$ 为第 0 天至 x 天的多日平均场与相应的月平均场的关系; 其他符号见正文

取平均的天数再多, 则其相关系数还不如只用第 0 天外推的相关。有些文献中定义的惯性预报为用前一个月的平均外推下一个月, 这种相关仅为 0.2—0.3, (见图 1 的外延) 远小于第 0 天外推的 0.58。Roads^[2]分析 500 hPa 高度场时也得到类似的结果。应该指出的是, 用第 0 天外推未来 30 天月平均的相关虽然高于用前 30 天的平均作外推的相关系数, 但其均方根误差却较大 (文中省略了均方根误差的分布图)。这是很自然的, 因为某一天波动的振幅肯定要比 30 天平均起来大得多。当用第 0 天外推预报时, 除于某个适当的常数就很容易解决这一问题, 因为这时两者的相关系数不变, 而均方根误差变小。所以, 评价长期预报结果的好坏, 应与用第 0 天外推的惯性相关作比较。本文下面所说的惯性预报均指用第 0 天外推的预报。

下面讨论日平均场的不同波数对月平均场相关系数的贡献。相关系数作为波数的函数其定义如下:

$$\rho_n = \frac{\sum_{m=0}^n (x_{c_n}^m y_{c_n}^m + x_{s_n}^m y_{s_n}^m)}{\left[\sum_{m=0}^n (x_{c_n}^{m^2} + x_{s_n}^{m^2}) \right]^{1/2} \left[\sum_{m=0}^n (y_{c_n}^{m^2} + y_{s_n}^{m^2}) \right]^{1/2}}, \quad (2)$$

ρ_n 为相关经向谱。(2) 式对 m 求和若改为对 n 求和, 则为相关纬向谱的定义^[3]。其中 x 和 y 为要考虑的两个场, 上、下标的含义与 (1) 式的相同。图 2 为某日平均场与月平均场相关的纬向谱 (图 2a) 和二维谱 (图 2b) 的分布。图中的“0”曲线为惯性预报。可见, 不同波数对惯性预报的贡献是不同的。不论是纬向谱还是二维谱, 分布都很不均匀, 除 $(n, m) = (0, 0)$ 外, $m = (2, 9, 11, 15)$ 有极值, $n = (2, 4)$ 的值也较大。图 2 中的曲线“-7”、“-3”、“+3”和“+7”分别表示前 8 天、前 4 天、第 3 天和第 7 天的日平均场与该月平均场的相关谱。可见, 随着时间的推移, 所以波数的相关不断增强, 且分布的形式比较相似。

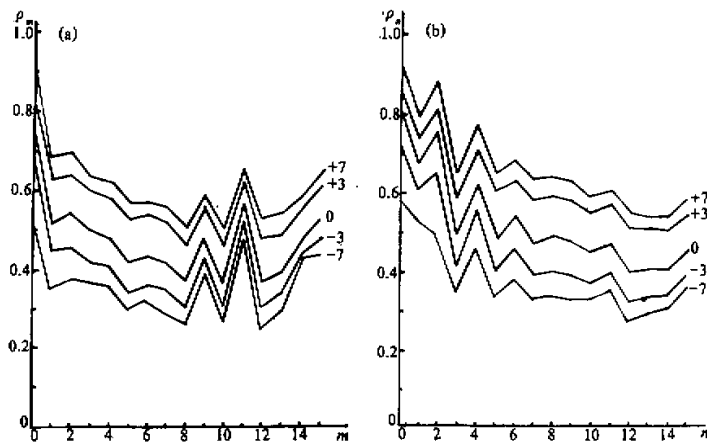


图 2 某日平均场与月平均场的相关纬向谱 (a) 和相关二维谱 (b)

说明见正文

四、月平均预报的几种统计模式

本节提出如下几种月预报的统计方法, 并限制只能用第 0 天以前的资料。

1. 用不同时间的日平均作预报因子

假设月平均满足下列回归方程:

$$\hat{T} = \sum_{i=-N}^0 a_i T(i), \quad (3)$$

其中 $T(i)$ 为第 i 天 1000 hPa 温度距平, $\hat{T}$ 为第 0 天后 30 天的平均, a_i 为回归系数, N 为天数, 可作适当选取。(3) 式也可看成对不同时间的日平均进行加权平均作未来 30 天的平均月预报。

把 15 年的资料分成前 8 年和后 7 年两组。第一组作非独立样本, 用最小二乘法拟合 $\hat{T}$, 求得 a_i 后, 对第二组资料作预报 (即独立样本)。然后反过来用第二组资料作非独立样本, 对第一组资料作预报。这样可得到 15 年的独立样本。取 $N=3$ 时, 两组的 a_i 值差别不大, 以第 0 天占的权重最大, 为 0.274, 其他 3 天的 a_i 值小得多, 约 0.05 左右。用它们代入 (3) 式, 得 15 年独立样本预报的相关系数的平均值为 0.61, 并用“x”标在图 1 中。它比惯性预报的 0.58 有所改进。若 N 取 5, 相关系数降至 0.59, N 取得再大, ρ 值降得越大。因此, 取前 4 天的日平均作回归因子能改善地面温度场的月预报, 不过改进不太大。

2. 用不同波数作预报因子

上节指出, 不同的波数具有不同的惯性, 因此, 也可作波数的加权处理。令

$$\hat{T} = a_n^m T(0), \quad (4)$$

其中, $T(0)$ 为第 0 天的 1000 hPa 温度距平的球谐系数, 它有 136 个非零值。(4) 式相当于用 136 个分量作预报因子的回归方程, a_n^m 为回归系数, 它也有 136 个值。同样把 15 年资料分两组, 得到两组的 a_n^m 不很均匀, 但都以 a_0^0 , a_4^0 最大, 为 0.65 左右。而且分布一般服从长波占的权重, 短波占权重小的规律。把 a_n^m 代入 (4) 式得到 15 年独立样本预报, 月预报的平均相关系数为 0.61, 优于惯性预报, 与用 (3) 式的回归方程结果差不多。

3. 用不同时间和波数作回归因子

把前两种方法结合起来, 令

$$\hat{T} = \sum_{i=-N}^0 a_n^m(i) T(i), \quad (5)$$

这是一个既用不同波数, 又用不同时间的日平均作因子的多元回归方程。以 $N=3$ 为

例, 权重 $a_n^m(i)$ 有 4×136 个。同样把 15 年资料分成两组, 最后得到 15 年独立样本的平均月预报的相关系数为 0.64, 用“ $\otimes$ ”符号标在图 1 中。它比 (3)、(4) 式的预报方程都好, 更比惯性预报好。

图 3 中给出了用 (5) 式作 1 天、10 天和 30 天平均预报的结果。曲线“W”为用 (5) 式预报的结果, “P”为惯性预报, “S”是用下一小节的统计-动力模式的结果。

4. 一个统计-动力预报模式

本节从动力学角度出发, 把统计方法和动力学方法结合起来, 并使统计模式的物理意义更清楚。

按热力学第一定律有^[4]

$$\frac{\partial T'}{\partial t} + LT' = Q', \quad (6)$$

其中, L 为一线性算子,

$$L = J(\bar{\psi}, T') + \frac{T'}{\tau_R}, \quad (7)$$

$\bar{\psi}$ 为 1000 hPa 的流函数的气候值, J 为 Jacobi 算子, τ_R 为牛顿冷却系数, Q' 是除了 LT' 项外所有对地表温度有贡献的因子, 如外界的加热率距平以及感热交换等。假设 Q' 可用经验公式表示成

$$Q' = \alpha_1 T', \quad (8)$$

代入 (6) 式后有

$$\frac{\partial T'}{\partial t} = \alpha T'. \quad (9)$$

这时, $\alpha = \alpha_1 - L$ 。 α 可用统计方法确定。这是一个源于动力学的统计模式, 可归为统计-动力学模式的一种。下面讨论 α 的确定方法以及推导平均预报公式。

(9) 式的前差近似为

$$\frac{T(j + \Delta t) - T(j)}{\Delta t} = \alpha T(j), \quad (10)$$

(上式的 T 略去了撇号)。 T 有 136 个非零值, 也就是说 (10) 式是 136 个方程组成的线性方程组。 α 的求法如下: 把变量 T 变成列向量形式, 将 (10) 式两端乘以 $T(j)$ 的转置, 稍加整理可得算子 α 的表达式:

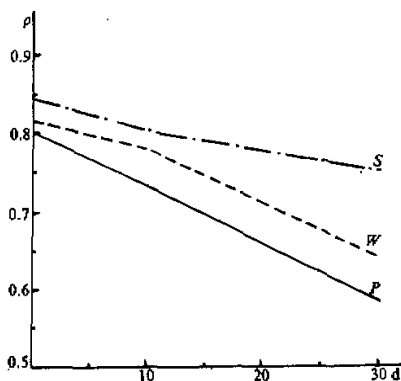


图 3 用不同预报方法作 1—30 天平均预报的结果

$$\alpha = \frac{1}{\Delta t} \{ \langle T(j + \Delta t) T^T(j) \rangle \langle T(j) T^T(j) \rangle^{-1} - I \}, \quad (11)$$

其中, 上标“ T ”和“ -1 ”分别为向量的转置和矩阵的逆, I 为单位矩阵, 两个角括号是所有非独立样本资料的相应平均值组成的落后协方差及协方差矩阵。 Δt 为资料的时间间隔, 这里是 12 小时, 每年的 j 从 0 到 245, α 是一个 136×136 的方阵。容易证明, 用 (11) 式求得的 α 与将 $T(0)$ 作为预报因子, $\partial T / \partial t$ 作为预报量, 用最小二乘法拟合 $\partial T / \partial t$ 的历史资料求回归系数 α_{jk} 的结果是一致的。也就是说, 按 (11) 式确定的 α 对非独立样本资料是最优的。(4) 式和 (9) 式除了预报量不同外 (前者为 T 的月平均, 后者为 T 的倾向), 前者 $\hat{T}$ 中之每一个只用其相同 m, n 的 $T(0)$ 拟合, 后者 $\partial T / \partial t$ 中的每一个要用所有的 m, n 的 $T(0)$ 拟合。其物理意义是不同波数之间有相互作用的。这样一来, 前者的回归系数为 136 个, 后者为 136×136 个。

由于未知数量很大, 要求非独立样本资料数足够多。为此, 不像以前简单地把资料分成两组而是采用交叉检验方法。比如, 图 1967—1980 年作非独立样本资料求出 α 作 1981 年独立样本预报; 用 1967—1979 以及 1981 年的资料求 α 作 1980 独立样本预报……依此类推, 同样可得 15 年的独立样本预报结果。

α 确定后, 容易从 (9) 式推导一个平均预报公式。(9) 式的积分解为

$$T(t_1) = T(0) + \alpha \int_0^{t_1} T(t) dt, \quad (12)$$

其中, t_1 为从零开始的某一时刻, $T(0)$ 为初值。如果要预报的是某一时段的平均天气, 则需再作一次积分:

$$\hat{T} = \frac{1}{T_2} \int_0^{T_2} T(t_1) dt_1 = T(0) + \frac{\alpha}{T_2} \int_0^{T_2} \int_0^{t_1} T(t) dt, \quad (13)$$

其中, T_2 为要预报的平均时段。再引入另一距平量:

$$T'' = T - \hat{T}, \quad (14)$$

T'' 为某一时刻的地面温度相对于气候值的距平与某一时段地面温度距平的平均值之差。由于 $\hat{T}$ 在 0 到 T_2 时段内是一个常数, 所以:

$$\int_0^{T_2} \int_0^{t_1} \hat{T} dt dt_1 = \hat{T} \int_0^{T_2} t_1 dt_1 = \hat{T} T_2^2 / 2,$$

把 (14) 式代入 (13) 式, 稍加整理有

$$\hat{T} = \left(1 - \frac{T_2}{2} \alpha \right)^{-1} \left\{ T(0) + \frac{\alpha}{T_2} \int_0^{T_1} \int_0^{t_1} T''(t) dt dt_1 \right\}, \quad (15)$$

类似的平均量公式最早由陈英仪等^[5]对正压涡度方程导出。由于 T'' 是减去气候平均值后再减去某一时段的距平平均值, 所以是一个 2 级小量, (15) 式的最后一项可以略去。即平均预报公式变成

$$\hat{T} = \left(1 - \frac{T_2}{2} \alpha\right)^{-1} T(0). \quad (16)$$

上式表明, 平均距平温度决定于初值 $T(0)$ 以及考虑了历史演变的算子 α 。用 (16) 式作 1—30 天平均预报的 15 年平均相关系数见图 3 的曲线“S”。它们大大地超过了惯性预报。当然, 对 1 天的预报而言, 统计—动力模式比不上好的大气环流模式。但对 10—30 天, 尤其是 30 天的长期预报, 现在的结果与实况的相关系数达 0.75, 均方根误差只有 1°C 左右 (图略)。还未看到数值预报有这么好的例子。

以上是 15 年的平均结果, 1967—1981 每年用统计—动力模式作的平均月预报结果见图 4。“S”曲线是模式的预报, “P”为惯性预报。可见, 每年预报也都优于惯性预报, 说明预报结果是稳定的, 不是偶然的。

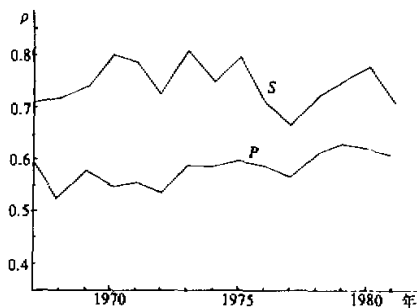


图 4 1967—1981 每年用统计—动力模式作的月预报结果 (“S”), 并与惯性预报 (“P”) 比较

五、结 论

本文用 15 年北半球夏季地面温度距平的资料, 分析了月平均场与该月前后一周左右的逐日平均场在时间和波数上的相关, 结果表明, 时间上越靠后的日平均场与月平均场的相关越强, 反之越弱。用第 0 天外推未来 30 天月预报的准确率比用前 30 天的平均作外推好。各种预报结果的评分标准应与前者定义的惯性预报作比较。

文中提出了几种月预报的统计模式。用前 4 天的日平均作预报因子以及用第 0 天的不同波数作预报因子的模式都能提高月预报的准确率。既用前 4 天的日平均, 又用这几天的不同波数作预报因子的模式更好。1395 个月预报例子的平均相关系数为 0.64, 高于惯性预报的 0.58。最好还是用一个统计—动力模式, 月平均预报的相关系数高达 0.75。这些模式预报试验的结果说明, 统计学方法还应是长期天气预报的重要手段, 值得做更多的试验并继续加以研究。

由于本文讨论的是夏季地面温度场, 它本身的持续性相关较高, 这也许是本模式预报效果好的原因。

参 考 文 献

- [1] Speth, P. and E. Kirk, 1981, Representation of meteorological fields by spherical harmonics, *Meteor. Rundsch.*, 34, 5—10.
- [2] Roads, J.Q. and T.P. Barnett, 1984, Forecasts of 500mb height using a dynamically oriented statistical model.

Mon. Wea. Rev., **112**, 1354—1369.

- [3] Roads, J.Q., 1986, Forecasts of time-averages with a numerical weather prediction model, *J. Atmos. Sci.*, **43**, 871—892.
- [4] Group of Long-Range Numerical Weather Forecasting, 1977, On the physical basis of a model of long-range numerical weather forecasting, *Scientia Sinica*, **20**, 377—390.
- [5] Chen, Y.Y., S. Chen and J.Q. Roads, 1991, Linear predictions and diagnosis of time-averages in a two-level model, *Tellus*, **43A**, 81—96.

Statistical Models for Monthly Surface Temperature Predictions in Summer Season

Chen Yingyi and Zhang Qiuqing

(National Research Center for Marine Environment Forecasts, Beijing 100081)

Abstract

The correlations between monthly mean and daily mean fields which are one week before or after that month have been analysed with 15-year record of the 1000 hPa temperature for the Northern Hemisphere. It is found that the newer the daily data are, the stronger the correlations. Persistent skills for 30-day averaged forecasts are much greater when using instantaneous daily data rather than ahead 30-day averages. It suggests that model skill measures should be compared to the persistence defined by the former.

Several statistical models are developed in this paper. Skills of all these models are better than those of persistence for monthly forecasts. A dynamically oriented statistical model is the best one. The mean correlation coefficient between monthly predictions with this model and the observations of about 1400 cases achieved 0.75 is higher than the persistence, 0.58.

Key words: long-range forecasts for surface temperature; statistical predictions; dynamic-statistical model.