

陆面模式 5 公里 3 小时大气驱动场及其扰动场的建立方案

大气驱动数据包括七个变量：近地面 2 米温度、地表压强、近地面 2 米相对湿度、近地面 10 米风速、降水率、下行短波辐射和下行长波辐射。建立陆地模式 5 公里 3 小时时间分辨率大气驱动数据的技术路线是：首先以薄板平滑样条函数和已存在的格点驱动资料为自变量，以台站观测资料为应变量，建立驱动场的趋势面；然后对趋势面的残差场进行订正；最后估计的订正场的误差协方差矩阵，并根据它随机生成已建立的驱动场的扰动场集合。

1 温度场

1.1 观测资料的处理

1958-1989 年的观测数据是日尺度的，每天台站观测个数大概在 670 个左右（图 1）。在这个时间段里，用日尺度的观测资料建立模型得到日尺度的 5 公里驱动场数据。然后用再分析资料进行 3 小时分配（见步骤 3）。

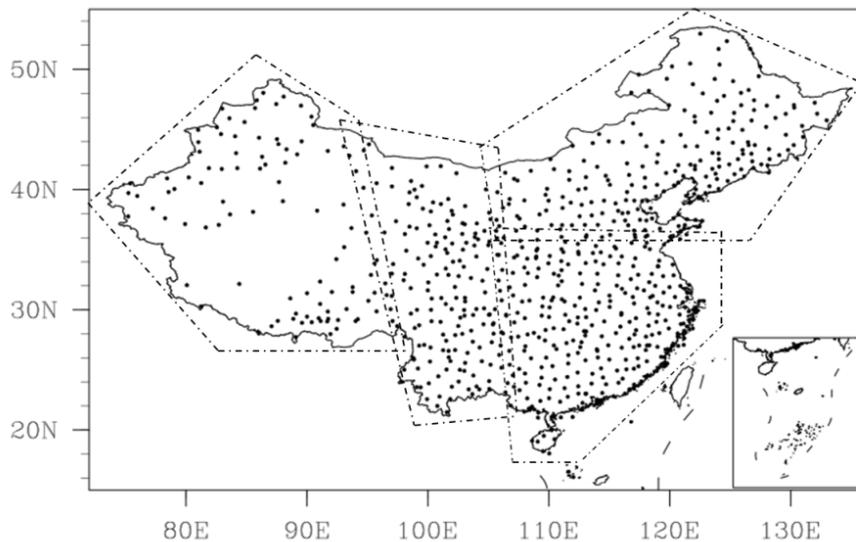


图 1. 项目所用的中国大陆区域观测台站的位置及子区域的划分

在 1990-2010 年的温度观测数据中，1990-1997 年是 6 小时分辨率的观测数据，这时可利用 6 小时数据对 3 小时数据的插值估计出 3 小时的观测数据；虽然 1998-2010 年的观测数据的时间分辨率是 3 小时，但是 1998-2006 年期间每天的 3、9、12、15 时的观测数据量比 0、6、12、18 四个时刻的观测数据量少很多，所以我们首先要将数据量少的这些时刻的观测数据补全（见本节的附 1）。通过以上方法，1990-2010 年的 3 小时观测数据就准备好了。

1.2 建立方案

1.2.1 利用薄板平滑样条模型建立趋势面

$$t(x, y) = f(x, y) + \alpha \cdot z(x, y) + \varepsilon(x, y), \quad (1)$$

其中 x 为经度, y 为纬度, $t(x, y)$ 为对应位置温度的真值, $z(x, y)$ 为对应位置的高程, $f(x, y)$ 为薄板平滑样条函数, α 为待估计的线性回归量系数, $\varepsilon(x, y)$ 为均值为 0, 方差为 σ^2 (空间统计独立的) 的误差变量。对于每一个时刻, 利用观测数据和经度纬度高程信息训练模型 Eq.(1) 从而估计出 $f(x, y)$ 的各项系数和 α 值, 以及 $\varepsilon(x, y)$ 的方差 σ^2 。将 5 公里格点的经纬度和高程信息输入模型 Eq.(1) 就得到趋势面, 即 $t_{trend}(x, y) = f(x, y) + \alpha \cdot z(x, y)$ 。上述过程可以通过 R 软件“gam”程序包中的“mgcv”和“predict”两个函数实现。

为了加快计算速度, 中国大陆区域被划分为四个边界重叠的子区域: W(西部), Mw(中西部), Ne(北东部), Se(南东部) (图 1)。先在每一个子区域上建立趋势面, 再将边界重叠的部分按到两边界距离加权平均, 以形成整个中国大陆区域的趋势面。

1.2.2 趋势面的订正

利用平滑样条方法建立的趋势面可能在有观测值的地方与观测的离差值远远大于观测误差, 所以我们用观测对趋势面进行订正。订正后的场就是我们估计的驱动场的温度场。具体的订正方法如下:

$$t_{adj}(x, y) \equiv t_{trend}(x, y) + \sum_{i=1}^{m(x,y)} w^2(x_i, y_i) \times (t_{obs}(x_i, y_i) - t_{trend}(x_i, y_i)) \bigg/ \sum_{j=1}^{m(x,y)} w(x_i, y_i), \quad (2)$$

其中 (x, y) 为待订正位置的经纬度, (x_i, y_i) 是观测点的经纬度; $m(x, y)$ 为以 (x, y) 为中心, 预先选定的值 R 为半径的圆内观测测点的个数; 此外,

$$w(x_i, y_i) = \frac{R^2 - d_i^2}{R^2 + d_i^2}, \quad (3)$$

其中 $d_i^2 = (x_i - x)^2 + (y_i - y)^2$ 。

根据平滑样条的假设, Eq. (1) 中的误差 $\varepsilon(x, y)$ 是空间统计独立的, 所以从理论上残差订正不会改进观测点上的估计。但在项目的执行过程中, 我们发现误差 $\varepsilon(x, y)$ 在实际例子中是空间相关的, 因此订正是必要的。订正不但可以使得温度场产品在观测附近更为平滑, 而且还可以提高温度场产品在观测稀疏地区的预报精度 (详见 5 结果验证)。

1.2.3 扰动场的建立

扰动场的建立分为两步。首先估计已建立的驱动场的误差方差矩阵（具体的公式推导见附 2），然后根据它生成零均值的多元正态随机向量。最后将生成的随机向量叠加在已建立的驱动场上便形成了驱动场的扰动场。

1.3 时间尺度上的分配

由于 1958-1989 年的温度只有日观测数据，所以得到的驱动场数据也是日尺度的。若要得到分辨率为 3 小时的驱动数据，需借助于再分析资料进行小时尺度上的分配。1958-1978 年我们采用 Princeton 再分析资料(附 3)进行分配，1979-1989 年采用 CFSR（NCEP Climate Forecast System Reanalysis）再分析资料（附 3）进行分配。具体分配的方法如下：

$$t_{3\text{hourly}} = t_{3\text{hourly}}^{\text{reanalyze}} - \bar{t}_{3\text{hourly}}^{\text{reanalyze}} + t_{\text{daily}}, \quad (4)$$

其中 $t_{3\text{hourly}}^{\text{reanalyze}}$ 是由再分析资料插值得到的 3 小时 5 公里分辨率的数据（附 3），此处 1958-1978 年应用的是 Princeton 再分析资料，1979-1989 年应用 CFSR 再分析资料， $\bar{t}_{3\text{hourly}}^{\text{reanalyze}}$ 为由再分析资料插值得到的 3 小时 5 公里格点处的日平均， t_{daily} 为用 1.2 节方法建立的日的温度的 5 公里格点场。

1.4 验证方法

2003 年全年的数据被选用来对各种方案都做检验。具体程序如下：

应用交叉叠合（Cross-validation, CV）的方法来检验不同的插值方法在台站集合上的效果。具体的程序是：对于每个时刻，将观测值 $U(x_n, y_n)$ ($n=1, 2, \dots, N$) 作为一个验证数据从观测中去除，用剩下的样本值训练模型，然后用训练好的模型预报 (x_n, y_n) 点的值（即 $\hat{U}(x_n, y_n)$ ）。该方法在同一时刻应用于所有的台站观测。评价统计量选用 CV 的定义式为：

$$CV = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{U}(x_n, y_n) - U(x_n, y_n))^2}, \quad (5)$$

其中， N 是台站数目。插值方案对应的 CV 值越小，意味着插值效果越好。

为评估我们的产品和其它空间分辨率的格点上产品（例如 CFSR 再分析资料 Climate Forecast System Reanalysis，空间分辨 0.3125° ）。比较方法如下：

首先，将我们的 5 公里驱动数据产品插值到 $0.05^\circ \times 0.05^\circ$ 的格点上，然后利用得到的 $0.05^\circ \times 0.05^\circ$ 分辨率的数据集平均得到 $0.3125^\circ \times 0.3125^\circ$ 分辨率的数据（该分辨率是 CFSR 数据的分辨率）。最后，在 $0.3125^\circ \times 0.3125^\circ$ 的分辨率上，两种产品在品格点上的差的方差估计为

$$\text{var}(U^{CFSR} - U^{BNU}) = \frac{1}{g \times i} \sum_g \sum_i (U_i^{CFSR}(g) - U_i^{BNU}(g))^2, \quad (6)$$

这里 U^{CFSR} 是 CFSR 数据， U^{BNU} 是我们的产品在 0.3125° 分辨率格点上的估计值， g 是格点数， i 是时间步数。假设 $\text{cov}(U^{CFSR}, U^{BNU}) = 0$ ，即 U^{CFSR} 和 U^{BNU} 的误差是统计独立的。可以证明(附 4) U^{CFSR} 的方差与方差 U^{BNU} 之差的下界为

$$\text{var}(U^{CFSR}) - \text{var}(U^{BNU}) \geq \text{var}(U^{CFSR} - U^{BNU}) - 2CV. \quad (7)$$

如果这个差是正的，就说明我们的产品在 CFSR 的格点上比 CFSR 产品的方差小。

1.5 结果验证

表 1 给出了我们 (BNU) 拟合的近地面 2 米温度场在观测台站 CV 值 ($^\circ\text{C}$) 和 Princeton 线性插值到观测台站的均方根误差 (RMSE) ($^\circ\text{C}$)。其中，温度的 Princeton 数据插值到观测台站的方法如下：先把 Princeton 数据按照 $0.0065^\circ\text{C}/\text{m}$ 的 lapse rate 降到海平面，然后用降高程的气温值进行线性插值到观测台站上，最后在用该 lapse rate 升回地形高度。表 1 中结果说明我们的产品显著地好于 Princeton 驱动资料。

表 1. 地表 2 米温度场在观测台站上的计算误差 ($^\circ\text{C}$).

	W	Mw	Ne	Se	All
BNU (CV)	2.89	2.02	1.79	1.35	2.01
Princeton (RMSE)	4.06	3.31	3.50	2.71	3.39

表 2 给出了 Princeton 驱动数据和 BNU 产品方差的差异的下界的平方根 ($^\circ\text{C}$, Eq.(7))。这些值都是正的，平均值在 3.1°C 左右，说明我们的产品显著地好于 Princeton 驱动数据。

表 2. Princeton 驱动数据方差和 BNU 温度产品方差的差异的下界平方根 $^\circ\text{C}$

W	Mw	Ne	Se	All
3.33	2.91	3.74	2.44	3.10

图 2、图 3 分别是 2003 年冬天和夏天两个温度场的分布。

The result of fitting t for year 2003 month 1 day 1 hour 0.

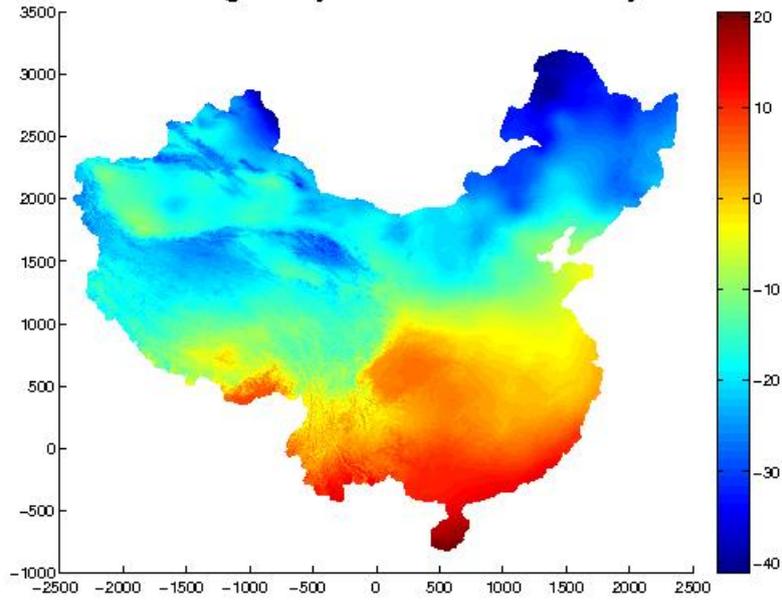


图 2. 2003 年 1 月 1 日 0 时温度场

The result of fitting t for year 2003 month 7 day 1 hour 0.

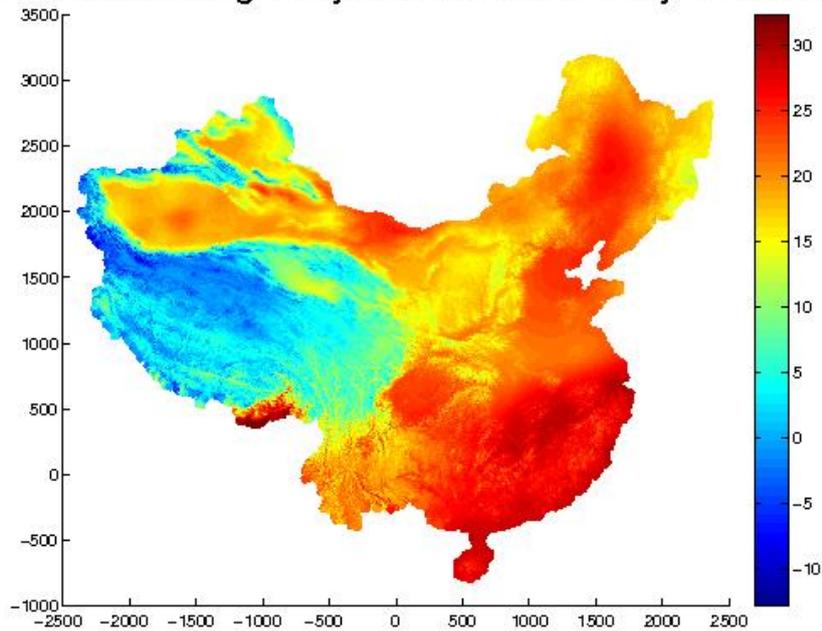


图 3. 2003 年 7 月 1 日 0 时温度场

Eq.(1)中的参数 α 可作为温度随高度下降的梯度的估计。图 4 显示了 2003 年四个子区域温度随高度下降的梯度的估计, 从图中可以看出每个子区域梯度都有明显的季节变化。

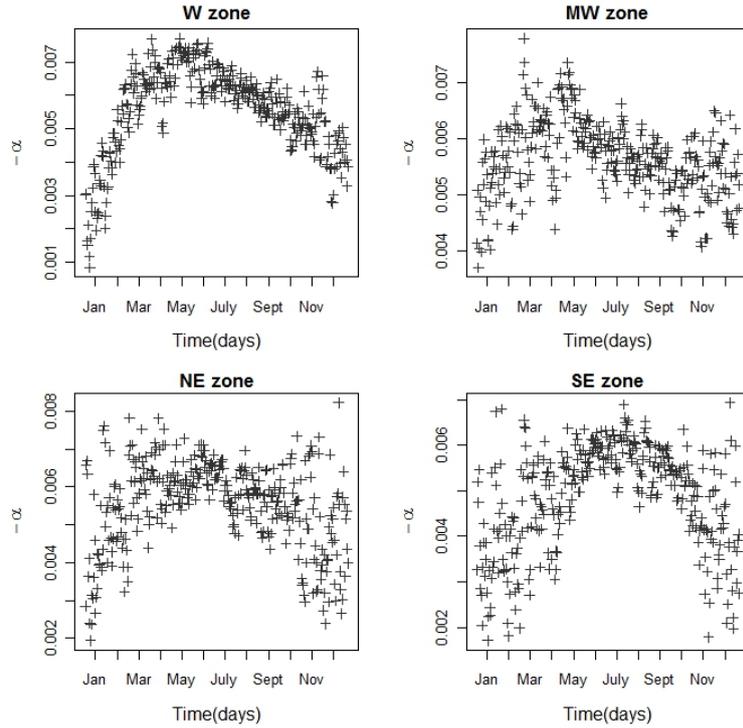


图 4. 四个子区域温度随高度下降的梯度

1.6 附件

附 1: 数据补差的方法

1998-2006 年虽然有三小时的观测数据, 只有 0 时、6 时、12 时、18 时的观测数据量是一致的, 而 3 时、9 时、15 时、21 时这四个时刻的观测数据数量一致, 且 3、9、15、21 时的观测数据量远少于 0、6、12、18 时的数据量。所以我们采用下述模型把 3、9、15、21 这四个时刻的观测数据补齐。补齐的具体程序是: 利用 i , $i-3$, $i+3$ 三个时刻共有的观测数据建立模型, 然后给定 i 时刻没有的那些观测台站第 $i-3$, $i+3$ 时刻的值及其经度纬度高程值就可以得到这些观测台站第 i 时刻估计的观测值。

$$t_i = f_i(x, y) + \alpha_i \cdot z(x, y) + \beta_i \cdot t_{i-3} + \gamma_i \cdot t_{i+3},$$

其中 i 代表时间, $f_i(x, y)$ 为薄板样条函数, $z(x, y)$ 为高程, α, β, γ 为待估计线性回归量系数。

附 2 订正场的误差协方差

由于观测误差几乎可以忽略, 我们假定对每一个观测点 x_j, y_j , 观测值等于真值, 即 $t(x_j, y_j) = t_{obs}(x_j, y_j)$ 。任意两点 x_1, y_1 和 x_2, y_2 间的趋势面的协方差为:

$$\begin{aligned} & \text{cov}(t_{trend}(x_1, y_1), t_{trend}(x_2, y_2)) \\ & \equiv E(t_{trend}(x_1, y_1) - t(x_1, y_1))(t_{trend}(x_2, y_2) - t(x_2, y_2)) \end{aligned}$$

$$= E(\varepsilon(x_1, y_1)\varepsilon(x_2, y_2)) = \sigma^2 c_{12},$$

这里， c_{12} 是 $\varepsilon(x_1, y_1)$ 和 $\varepsilon(x_2, y_2)$ 之间的相关系数，并假定它只与两点的距离 d_{ij} 有关，即 $c_{ij} = c(d_{ij})$ ，函数 $c(d)$ 可以通过 R 软件“spatial”程序包中的函数“correlogram”实现。

订正场中任意两点 x_1, y_1 和 x_2, y_2 之间的协方差为：

$$\begin{aligned} \text{cov}(t_{adj}(x_1, y_1), t_{adj}(x_2, y_2)) &\equiv E\left[\left(t_{adj}(x_1, y_1) - t(x_1, y_1)\right) \cdot \left(t_{adj}(x_2, y_2) - t(x_2, y_2)\right)\right] \\ &= E\left[\left(t_{trend}(x_1, y_1) - t(x_1, y_1) + \sum_{i=1}^{m(x,y)} \left(\frac{w_{1i}^2}{\sum_k w_{1k}} \cdot (t_{obs}(x_i, y_i) - t_{trend}(x_i, y_i))\right)\right) \cdot \right. \\ &\quad \left. \left(t_{trend}(x_2, y_2) - t(x_2, y_2) + \sum_{i=1}^{m(x,y)} \left(\frac{w_{1i}^2}{\sum_k w_{1k}} \cdot (t_{obs}(x_i, y_i) - t_{trend}(x_i, y_i))\right)\right)\right] \\ &= \sigma^2 \left[c_{12} - \sum_{j=1}^{m(x,y)} \left(\frac{w_{2j}^2 c_{1j}}{\sum_k w_{2k}}\right) - \sum_{i=1}^{m(x,y)} \left(\frac{w_{1i}^2 c_{2i}}{\sum_k w_{1k}}\right) + \sum_{i,j} \left(\frac{w_{1i}^2}{\sum_k w_{1k}} \cdot \frac{w_{2j}^2}{\sum_k w_{2k}} \cdot c_{ij}\right) \right], \end{aligned}$$

类似地，订正场在任意一点 x_0, y_0 的方差估计是

$$\text{var}(t_{ads}(x_0, y_0)) = \sigma^2 \left[1 - 2 \sum_{j=1}^{m(x_0, y_0)} \left(\frac{w_{0j}^2}{\sum_k w_{0k}} \cdot c_{0j}\right) + \sum_{i,j} \left(\frac{w_{0i}^2 \cdot w_{0j}^2}{(\sum_k w_{0k})^2} \cdot c_{i,j}\right) \right].$$

特别地，如果 (x_j, y_j) 是影响半径 R 中的唯一一点，订正场的估计方差为

$$\text{var}(p_a(x_0)) = \sigma_b^2 (1 - 2w_{0j}c_{0j} + w_{0j}^2).$$

附 3：再分析资料插值到 3 小时 5 公里的格点

Princeton 数据是为了驱动陆面水文模型而建立的一套全球 50 年的分辨率为 3 小时 1.0° 的气象驱动再分析数据集，Sheffield. J; CFSR 是一套 1979–2009 年全球范围的陆地分辨率为 0.3125° ，6 小时再分析资料。两类再分析资料与中国大陆的位置关系如图 5 所示，其中蓝色格点为再分析资料，红色格点为中国大陆。

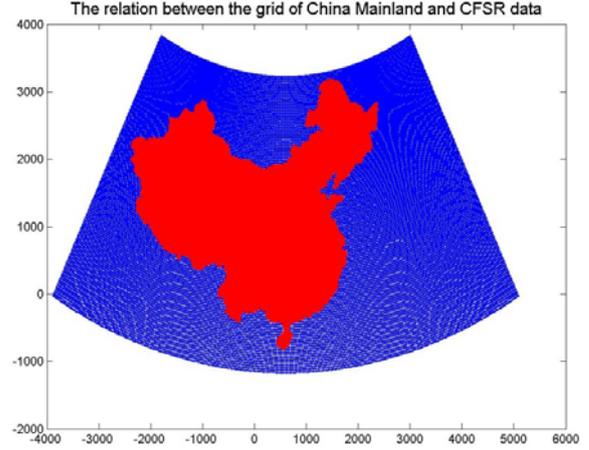
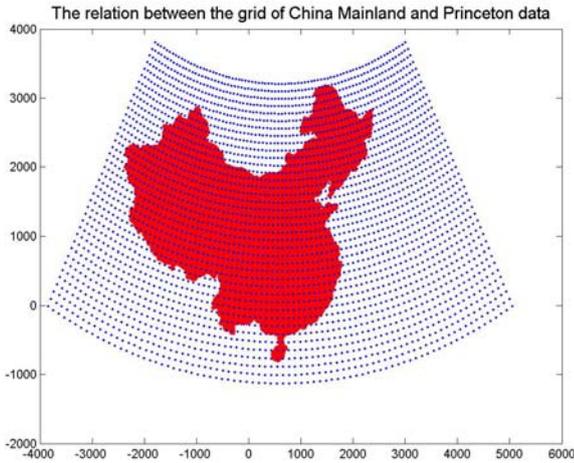


图 5. 中国大陆与再分析资料位置关系

如果是用的 Princeton 3 小时的数据，那么就只需要把它插值到 5km 的格点上就是 Eq.(4)中的 $t_{3hourly}^{reanalyze}$ ；如果是 CFSR 6 小时的再分析资料，那么先利用前后时刻的平均出 3 小时的数据，然后再把它插值到 5km 的格点就是 Eq.(4)中的 $t_{3hourly}^{reanalyze}$ 。

我们用 R 软件中多项式回归函数 loess 来解决这个问题，拟合程序为：

$$loess(t_{reanalyze} \sim x_{reanalyze} + y_{reanalyze}),$$

其中 $t_{reanalyze}$ 为 3 小时时间分辨率的温度再分析资料， $x_{reanalyze}$ 、 $y_{reanalyze}$ 为再分析资料的经纬度信息，利用这三个量建立模型，然后就可以根据 5 公里分辨率的格点经纬度信息预报出再分析资料插值到 5 公里的格点，得到格点的插值结果。

附 4 Eq.(7) 的证明

$$\begin{aligned} & \text{var}(U^{CFSR}) - \text{var}(U^{BNU}) \\ &= \text{var}(U^{CFSR} - U^{BNU}) - 2\text{var}(U^{BNU}) + 2\text{cov}(U^{CFSR}, U^{BNU}). \end{aligned}$$

利用 U^{CFSR} 和 U^{BNU} 的误差是统计独立的假设，我们得到

$$\text{var}(U^{CFSR}) - \text{var}(U^{BNU}) = \text{var}(U^{CFSR} - U^{BNU}) - 2\text{var}(U^{BNU}).$$

又由于 U^{BNU} 是我们的产品在 0.3125° 分辨率格点上的估计值，它的方差应不大于我们原始产品方差，即 $\text{var}(U^{BNU}) \leq CV$ ，于是得到 Eq.(7)。

2 气压场

2.1 观测资料的处理

该部分的处理方法同温度。

2.2 建立方案

2.2.1 利用薄板平滑样条模型建立趋势面

$$p(x, y) = f(x, y) + s(z(x, y)) + \varepsilon(x, y), \quad (8)$$

其中 $p(x, y)$ 为气压， $s(z(x, y))$ 为关于高程 $z(x, y)$ 的一维薄板样条函数，其它同 Eq.(1)。对于每一个时刻，利用气压台站观测数据和经度、纬度以及高程信息训练模型 Eq.(8)，估计出 $f(x, y)$ 和 $s(z(x, y))$ 的各项系数，然后只需给定某位置的经纬度和高程信息就可以通过建立的 Eq.(8) 预报得到 5 公里格点位置的拟合结果。

2.2.1 趋势面的订正

由于在观测台站附近的残差在观测误差的范围之内，趋势面不进行订正。

2.3 时间尺度上的分配

关于观测数据的处理和只有日观测资料时的分配法则，都和温度一致。

2.4 验证方法

与温度一致。

2.5 结果

表 3 给出了我们拟合的地表气压在观测台站 CV 值 ($^{\circ}\text{C}$) 和 Princeton 线性插值到观测台站的均方根误差 (RMSE) ($^{\circ}\text{C}$)。其中，气压的 Princeton 数据在台站上的插值方法如下：首先利用 Eq. (9) 把 Princeton 的气压数据根据压高公式降到海平面：

$$P_{SL} = P_H \left(1 + \frac{0.0065H}{T_H} \right)^{5.2568}, \quad (9)$$

其中 P_{SL} 为海平面气压， H 为高程值， P_H 为 H 高度 Princeton 数据的气压值， T_H 为地表气温；再把降高程的 Princeton 气压数据线性插值到台站上，最后按 Eq.(10)

把台站上的插值结果升回地形高度。

$$P_H = P_{SL} \left(1 + \frac{0.0065H}{T_H} \right)^{-5.2568} \quad (10)$$

表 3. 气压在观测台站上的计算误差 (hPa).

	W	Mw	Ne	Se	All
BNU (CV)	1.78	1.75	1.15	1.17	1.46
Princeton (RMSE)	4.89	4.03	3.84	3.68	4.11

表 4 给出了 Princeton 驱动数据和 BNU 气压产品方差的差异的下界的平方根 ($^{\circ}$ C, Eq.(7)). 这些值都是正的, 平均值在 4.2hPa 左右, 而且西部大于东部。说明我们的产品显著地好于 Princeton 驱动数据。

表 4. Princeton 驱动数据和 BNU 气压产品方差上的差异的下界平方根 (hPa).

区域	W	Mw	Ne	Se	All
Princeton	5.26	4.68	3.83	2.98	4.19

图 6 和

图 7 分别是 2003 年冬天和夏天两个气压场的分布。

The result of fitting p for year 2003 month 1 day 1 hour 0.

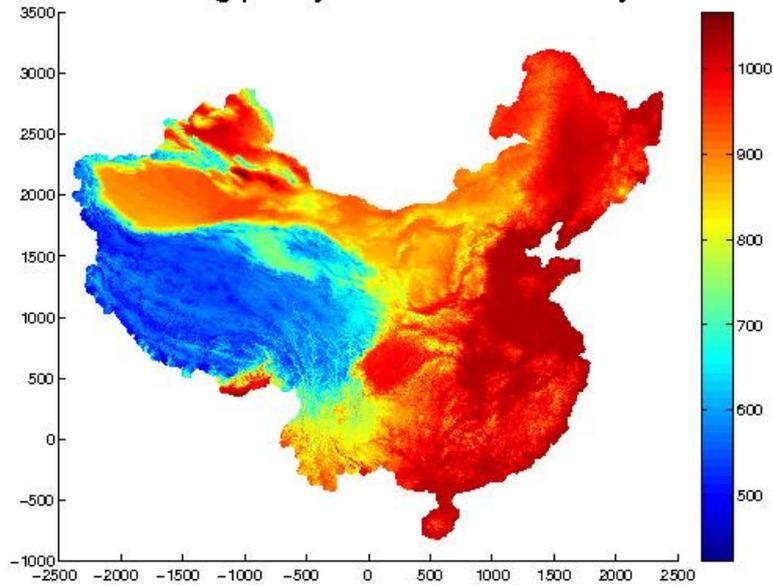


图 6. 2003 年 1 月 1 日 0 时气压场

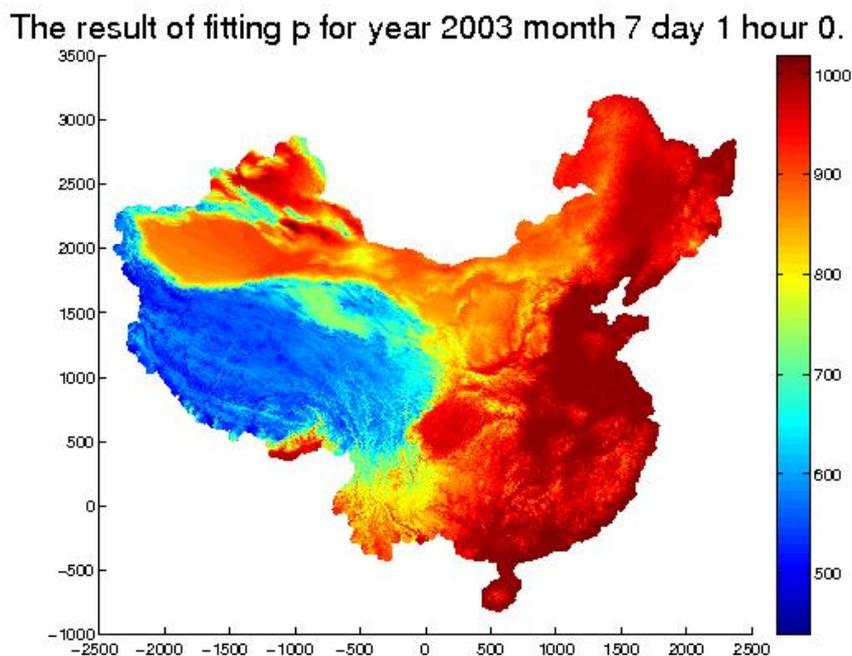


图 7. 2003 年 7 月 1 日 0 时气压场

3 相对湿度场的建立方案

3.1 观测资料的处理

观测资料的处理同温度。

3.2 建立方案

3.2.1 利用薄板平滑样条模型建立趋势面

1958–1978 年采用的模型为：

$$rh(x, y) = f(x, y) + \varepsilon(x, y), \quad (11)$$

1979–2010 年采用的模型为：

$$rh(x, y) = f(x, y) + \alpha \cdot rh_{CFSR}(x, y) + \varepsilon(x, y), \quad (12)$$

其中 $rh(x, y)$ 为相对湿度，其结果是由观测的比湿，气压和温度得到（附 5）。

$rh_{CFSR}(x, y)$ 是将 CFSR 的相对湿度值插值到 (x, y) 点的结果。

此外，为了保持所得数据的物理意义，我们将结果中大于 100 的部分变为 100，将结果中小于 0 的部分变为 0。

3.2.2 趋势面的订正

方法和温度的相同。

3.3 时间尺度上的分配

对于 1958–1989 年的相对湿度只有日观测数据，所以得到的驱动场数据也是日尺度的，想要得到小时的驱动数据就需要借助于再分析资料进行小时尺度上的分配。1958–1978 年我们采用 Princeton 数据（附 3）进行分配，1979–1989 年采用 CFSR 数据（附 3）进行分配。具体分配的方法如下：

$$rh_{3j_hour} = rh_{daily}^{obs} \cdot \frac{rh_{3j_hour}^{reanalyze}}{\sum_{i=0}^7 rh_{3i_hour}^{reanalyze}}, \quad (13)$$

其中， $rh_{3i_hour}^{reanalyze}$ 是由再分析资料插值得到的 3 小时 5 公里分辨率的资料（插值方式与温度场所用方法相同）， rh_{daily}^{obs} 为相对湿度的日观测拟合所得结果。对于计算得到的 rh_{3j_hour} 结果中大于 100 的部分，我们将其变为 100；小于 0 的部分，我们将其变为 0。

3.4 验证方法

与温度一致。

3.5 结果

表 5 给出了我们拟合的相对湿度在观测台站 CV 值(%) 和 Princeton 数据插值到观测台站得到 RMSE (%)。其中，Princeton 数据插值到台站得到 RMSE 的方法：直接线性插值。表 5 说明我们的产品要好于 Princeton 产品。

表 5. 相对湿度场在观测台站上的计算误差 (%)

区域	W	Mw	Ne	Se	All
BNU (CV)	14.07	11.40	10.03	8.75	11.06
Princeton (RMSR)	19.69	16.40	16.10	12.81	16.25

表 6 给出了 Princeton 驱动数据和 BNU 相对湿度产品方差的差异的下界的平方根(%)。这些值都是正的，平均值在 11.6%左右，而且西部地区差异尤其大，说明我们的产品显著地好于 Princeton 驱动数据。

表 6. Princeton 驱动数据方差和 BNU 产品方差上的差异的下界平方根 (%)

区域	W	Mw	Ne	Se	All
----	---	----	----	----	-----

Princeton	12.11	12.36	14.38	7.80	11.66
-----------	-------	-------	-------	------	-------

图 8 和图 9 分别是 2003 年冬天和夏天两个相对湿度场的分布。

The result of fitting rh for year 2003 month 1 day 1 hour 0.

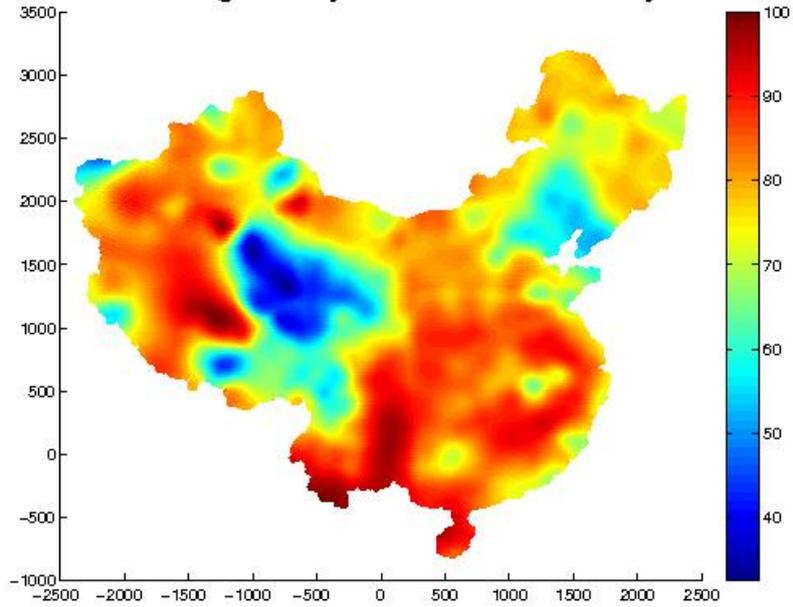


图 8. 2003 年 1 月 1 日 0 时相对湿度场

The result of fitting rh for year 2003 month 7 day 1 hour 0.

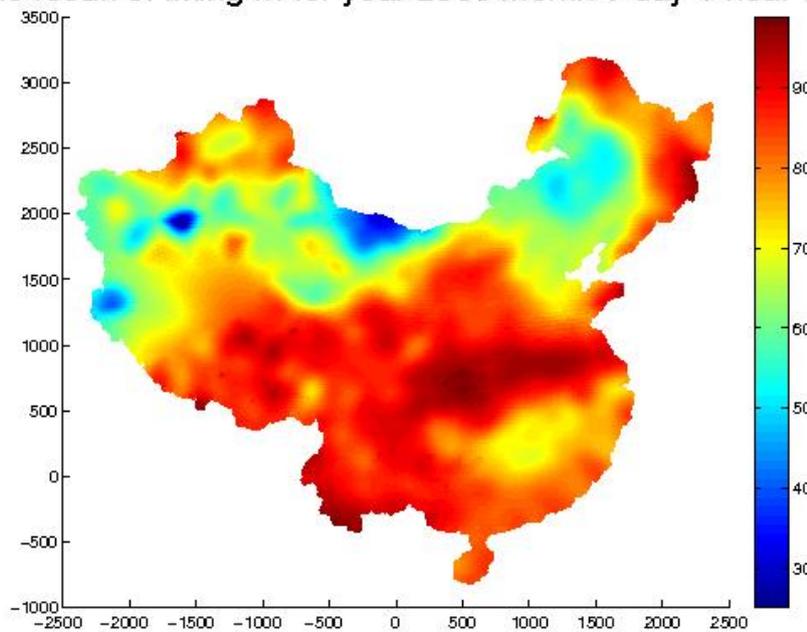


图 9. 2003 年 7 月 1 日 0 时相对湿度场

3.6 附件

附 5: 比湿转换成相对湿度:

比湿转换为相对湿度需要用到同时刻的气压和温度的值, 转换公式如下:

$$rh = \frac{q \times p \times 10^{\frac{-7.45 \cdot t}{235+t}}}{0.622 \times 6.11} \times 100,$$

其中 q 为比湿, p 为气压, 单位为百帕, t 为温度, 单位为 $^{\circ}\text{C}$ 。

4 风场

4.1 观测资料的处理

观测资料为日时间尺度。

4.2 建立方案

4.2.1 利用薄板平滑样条模型建立趋势面

1958–1978 和 2010 年采用的模型为:

$$w(x, y) = f(x, y) + \varepsilon(x, y), \quad (14)$$

其中 $w(x, y)$ 为经纬度 (x, y) 的风值。

1979–2009 年采用的模型为:

$$w(x, y) = f(x, y) + w_{\text{CFRS}}(x, y) + \varepsilon(x, y), \quad (15)$$

其中 $w_{\text{CFRS}}(x, y)$ 为 CFRS 风速在 (x, y) 处的插值。

4.2.2 趋势面的订正

方法和温度场的相同。

4.3 时间尺度上的分配

1958–1978 年: 采用 Princeton 数据分配。

1979–2009 年: 采用 CFRS 数据分配。

2010 年采用 NCEP 再分析资料 (附 6)。

4.4 验证方法

与温度一致。

4.5 结果

表 7 给出了我们拟合的风速场在观测台站 CV 值 (m/s) 和 CFSR 数据在观测站点上插值的均方根误差 (m/s)。这些值明显大于表 7 中的值, 说明我们的产品在小空间尺度上显著地好于 CFSR 数据。

表 7. 风速场在观测台站上的计算误差 (m/s)

	W	Mw	Ne	Se
BNU(CV)	1.52	0.99	1.12	1.18
CFSR (RMSE)	2.16	1.73	1.73	1.51

图 10 和图 11 分别 2003 年是冬天和夏天两个风速场的分布。

The result of fitting w for year 2003 month 1 day 1 hour 0.

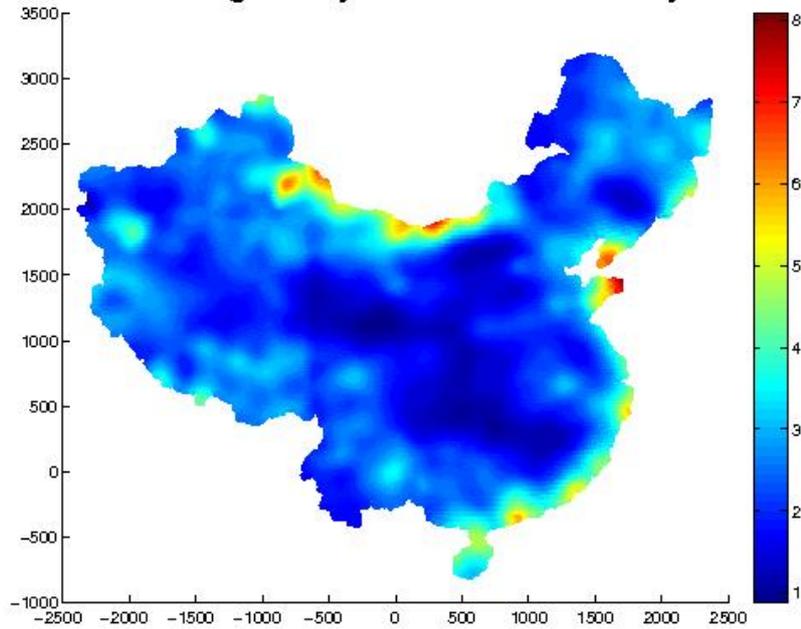


图 10. 2003 年 1 月 1 日 0 时风速场

The result of fitting w for year 2003 month 7 day 1 hour 0.

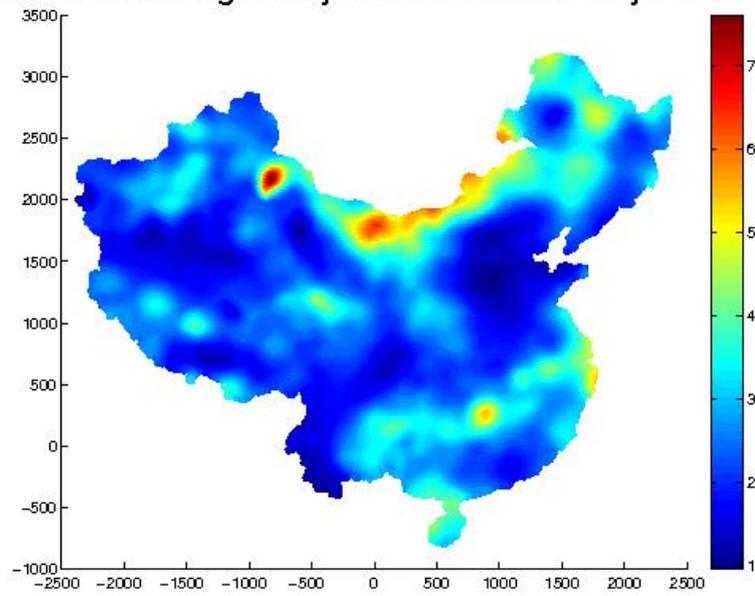


图 11. 2003 年 7 月 1 日 0 时风速场

4.6 附件

附 6： 2010 年 NCEP 再分析资料

2010 年 NCEP 资料与中国大陆的位置关系如图 12 所示：

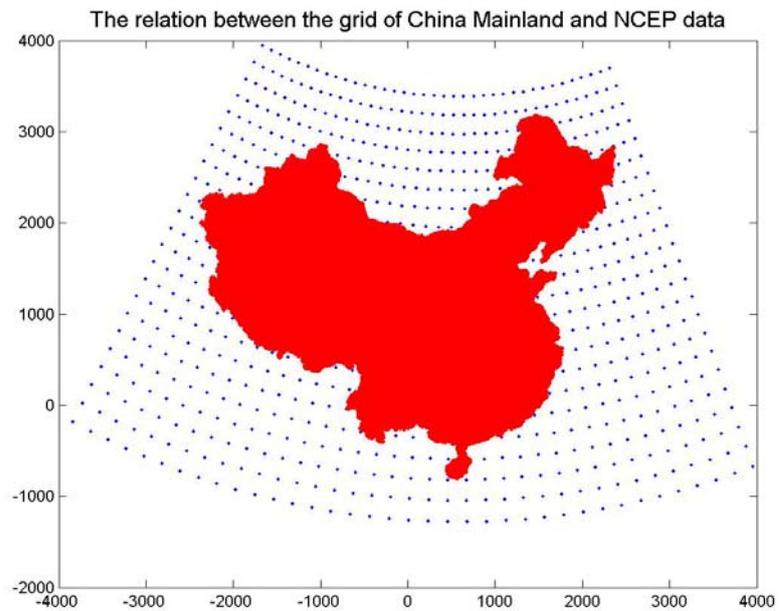


图 12. 中国大陆与 2010 年 NCEP 再分析资料位置关系
插值方法与 CFSR 数据相同。

5 降水

5.1 观测资料的处理

观测资料为日时间尺度。

5.2 建立方案

5.2.1 利用薄板平滑样条模型建立趋势面

1958–1978 年年采用的模型为：

$$r(x, y) = f(x, y) + \varepsilon(x, y), \quad (16)$$

其中 $r(x, y)$ 为经纬度 (x, y) 的降水值。

1979–1997 年年采用的模型为：

$$r(x, y) = f(x, y) + \alpha \cdot r_{CFSR}(x, y) + \varepsilon(x, y), \quad (17)$$

其中 $r(x, y)$ 为经纬度 (x, y) 的降水值， $r_{CFSR}(x, y)$ 为日平均 CFSR 降水产品在 (x, y) 的插值。

1998–2010 年年采用的模型为：

$$r(x, y) = f(x, y) + \alpha \cdot r_{CMORPH}(x, y) + \varepsilon(x, y), \quad (18)$$

其中 $r(x, y)$ 为经纬度 (x, y) 的降水值， $r_{CMORPH}(x, y)$ 为日平均 CMORPH (CPC MORPHing technique) 降水产品在 (x, y) 的插值。

5.2.2 趋势面的订正

方法和温度的相同。

5.3 时间尺度上的分配

1958–1978 年用 Princeton 资料分配。

1979–1997 年用 CFSR 资料分配。

1998–2010 用 CMORPH 资料分配。

5.4 验证方法

与温度一致。

5.5 结果

由于 700 个左右的台站的观测资料不能够有效地反映出中国大陆地区降水

的空间变化，我们只将 BNU 产品的日平均与 CMORPH 产品比较。在中国大陆地区，BNU 和 CMORPH 降水产品日平均的 CV 分别是 4.34mm/d 和 6mm/d。可见 BNU 降水产品还是由于 CMORPH 降水产品。

图 13 和图 14 分别是 2003 年冬天和夏天两个降水场的分布。

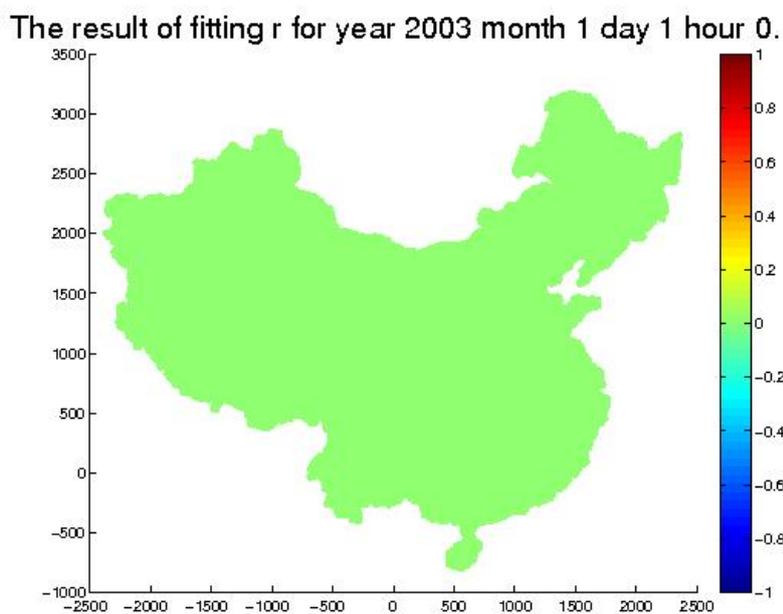


图 13. 2003 年 1 月 1 日 0 时降水场

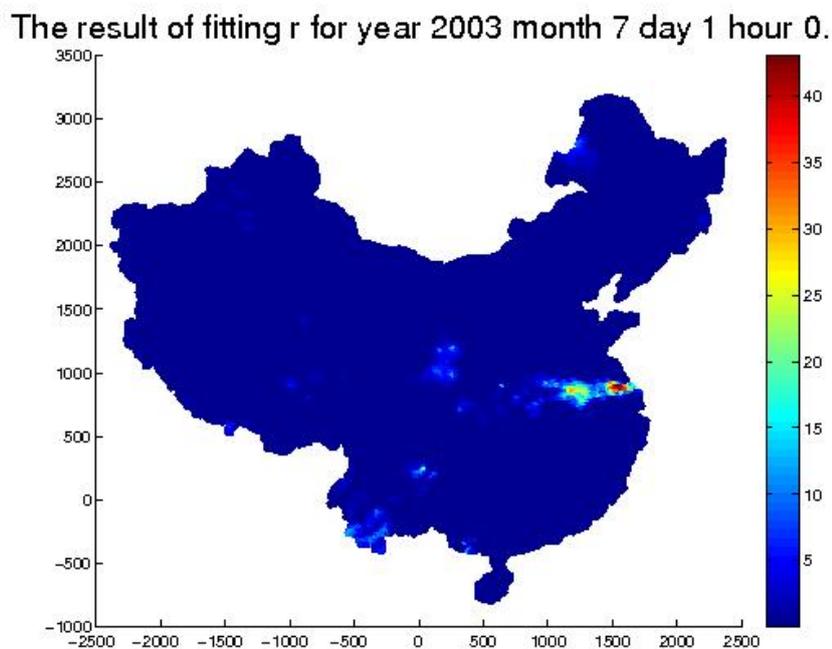


图 14. 2003 年 7 月 1 日 0 时降水场

6 下行短波辐射

6.1 观测资料的处理

日照时数的观测数据是日值，记录了每天的太阳垂直直接辐射强度大于 120W 的时间长度；其他所需资料为日最高气温、日最低气温和日均相对湿度。

6.2 模型的建立

首先，利用下列薄板平滑样条建立日照时数的 5 公里插值场（空间插值），

$$s(x, y) = f(x, y) + \beta t_d(x, y) + \gamma F(rh(x, y)) + \mu G(x, y) + \varepsilon(x, y), \quad (19)$$

其中 x 为经度， y 为纬度， t_d 为温度日变化，即日最高气温减去日最低气温， $f(x, y)$ 为薄板平滑样条函数， β ， γ ， μ 为线性回归量系数， $\varepsilon(x, y)$ 为均值为 0 的模型误差， rh 为日均相对湿度，它的函数 $F(rh)$ 为

$$F(rh) = \frac{1 - \exp(-2.34(1 - \frac{rh}{100}))}{1 - \exp(-2.34)}, \quad (20)$$

$G(x, y)$ 是 GEWEX SRB (Global Energy and Water Cycle Experiment Surface Radiation Budget) 是短波辐射遥感产品在 (x, y) 处的插值。

插值日照时数时，中国大陆区域分区方案同其他变量。建立模型时，温度和相对湿度日值数据采用观测值，而在格点上预报时，采用已经生成的 5 公里 3 小时格点场来计算。

然后，利用中国科学院青藏高原研究所阳坤教授等所提出的太阳辐射模型计算 5 公里 3 小时的下行短波辐射。模型参考文献如下：

Yang, K., Huang, G. W. & Tamai, N. (2001). *A hybrid model for estimating global solar radiation*. *Solar Energy*, 70(1), 13-22.

Yang, K., Koike, T. & Ye, B. (2006). *Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets*. *Agricultural and Forest Meteorology*, 137(1-2), 43-55.

具体计算过程为：

(1) 根据温度、相对湿度、气压的 5 公里 3 小时插值场，计算北京时间下的日均温度、日均相对湿度和日均气压，结合已经生成的日照时数插值场，计算下行短波辐射的日值插值场；

(2) 根据当地太阳高度角的变化，将下行短波辐射日值分配到 3 小时；

(3) 将北京时间换算到世界时。

GEWEX SRB (Global Energy and Water Cycle Experiment Surface Radiation Budget) 是长短波辐射遥感产品，数据自 1983 年 7 月开始，经验证质量比较理想。在 1983 年 7 月以后的日照时数插值中，将其中的短波辐射作为第三个协变

量来建立模型和做预报。

6.3 验证方法

下行短波辐射有日值观测数据，但仅有不足 90 个观测站点，我们使用这些日值观测作为“真值”来验证产品质量。同时对比 Princeton 再分析资料和 GEWEX SRB 遥感反演产品，2003 年验证结果见表 9。

表 9. 下行短波辐射精度验证 (W/m^2)

区域	W	Mw	Ne	Se	All
Princeton	51.392	58.093	54.939	62.134	57.653
GEWEX	31.452	34.671	29.324	30.330	31.408
BNU(t 和 rh 做协变量)	34.089	37.258	28.806	35.720	34.435
BNU(t、rh 和 GEWEX 数据做协变量)	28.308	33.646	25.700	32.934	30.868

结果表明，5 公里 3 小时插值产品的精度大大优于 Princeton 再分析资料。未使用 GEWEX 资料做协变量时，精度与 GEWEX 数据相当，而加入了 GEWEX 数据以后，精度要好于 GEWEX 数据。因此我们可以期望，在没有遥感产品的年份，产品质量能够大抵相当于 GEWEX 的质量；而在有 GEWEX 产品的年份，通过融合，可以得到优于 GEWEX 产品的数据场。

图 15 和图 16 分别是 2003 年冬天和夏天两个下行短波辐射场的分布。

The result of fitting srad for year 2003 month 1 day 1 hour 0.

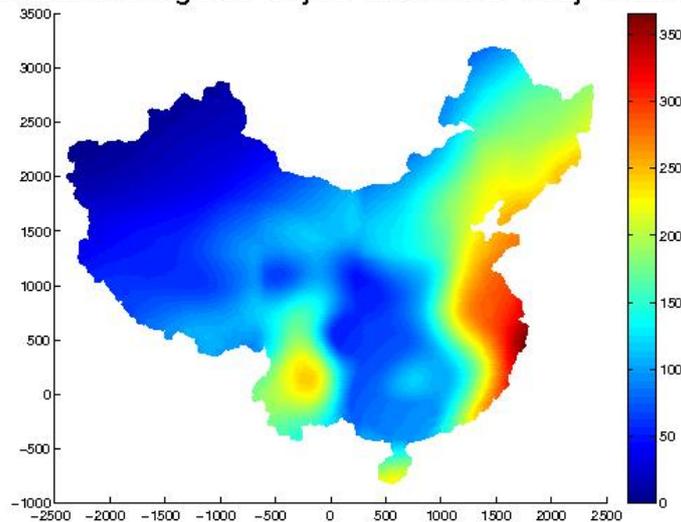


图 15. 2003 年 1 月 1 日 0 时下行短波辐射场

The result of fitting srad for year 2003 month 7 day 1 hour 0.

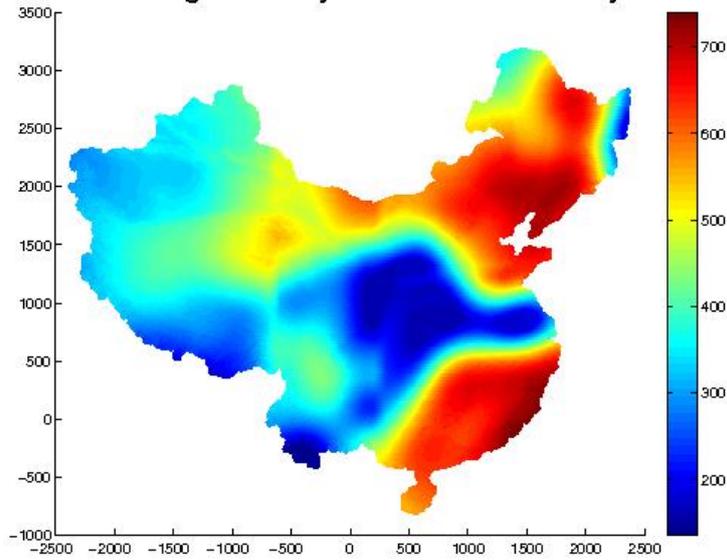


图 16. 2003 年 1 月 1 日 0 时下行短波辐射场

7 下行长波辐射

7.1 模型建立

长波辐射的值由短波辐射、温度、相对湿度和气压的 5 公里 3 小时插值场直接计算生成，模型来自文献：

Crawford, Todd M., Claude E. Duchon, 1999: *An Improved Parameterization for Estimating Effective Atmospheric Emissivity for Use in Calculating Daytime Downwelling Longwave Radiation*. J. Appl. Meteor., 38, 474–480.

7.2 结果验证

由于长波辐射无观测值，因此暂无科学的验证方法。

图 17 和图 18 分别是 2003 年冬天和夏天两个下行长波辐射场的分布。

The result of fitting lrad for year 2003 month 1 day 1 hour 0.

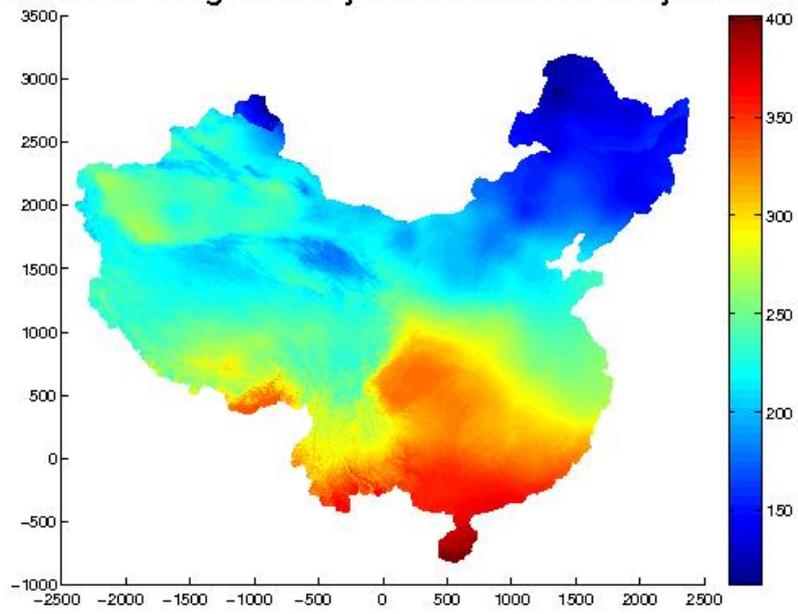


图 17. 2003 年 1 月 1 日 0 时下行长波辐射场

The result of fitting lrad for year 2003 month 7 day 1 hour 0.

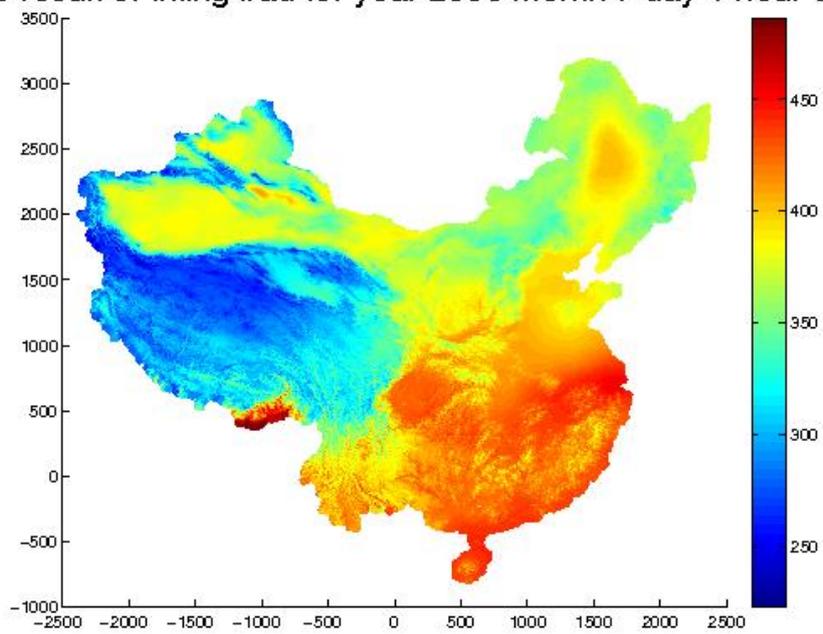


图 18. 2003 年 7 月 1 日 0 时下行长波辐射场