多个“平稳自回归序列”模型的综合予报

吴 荣
（南开大学）

本文研究的问题是：我国东部地区（华北、华东、华南等地），$M_S \geqslant 5.5$级地震的发震时间、地点的预报问题。

基本思想及资料处理

东部地区的发震问题不仅与本身有关，而且也与周围其它地区的发震情况有关。为此，取1900年—1976年，下列地区的历史资料进行研究，即：东部$M_S \geqslant 5.5$级震，西部$M_S \geqslant 6$级，台湾省$M_S \geqslant 6$级，国外：阿留申、千岛、日本、菲律宾、印尼、缅甸、印度、苏联、蒙古、阿富汗、直到土耳其、希腊等国$M_S \geqslant 7$级震（余震只取三个月以外较大者，如唐山余震，只取三个月以外$M_S \geqslant 6$级左右或以上的。1977年的资料，留作后验。

在发震时间的研究上，过去往往着重于研究某地区（如日本）发震后，对下一次东部发震时间的直接影响，这种时间间隔序列，本文中称为震后序列。除研究这种序列外，还着重研究其间接影响，即某地区发震后与紧随的第二次东部震的时间间隔，称为震前序列。在实际预报中，震前序列往往显得更重要，因为，当东部发震后，要预报下一次东部震何时发生，此刻震后序列尚未出现，而震前序列却已全部出现。自然，为了使预报时段缩短并补充震前序列之不足，震后序列的研究也是不可缺少的。文中的一些例子将说明这两种序列在预报中所起的作用。

为了用比较恰当的数学模型来处理，必须打破国土的界线，根据地质构造的特点，以及这些序列本身在不同地区呈现的不同规律，进行划区，调整为九个大区（东部本身构成一个区，称为第九区）。对每一个区都研究了前、后两种时间序列，然后再将它们综合起来，给出一个时间预报的模式。关于地点（将东部地区，按30°N为界，分成南、北两部分）。我们用概率回归估计法，作了两套公式，进行预报。

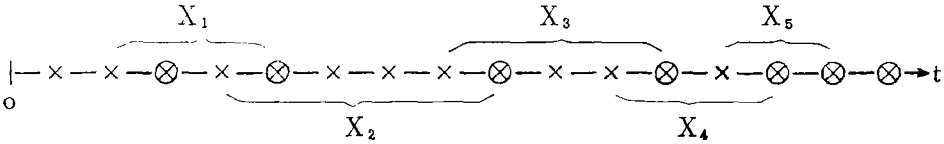
数学模型的建立

对所选取的每一个地区（第九区除外），取出前后两种时间序列，下面以某第i区为例，

参加这项计算工作的还有鲍淑萍、滦云风、王维志和杨润志。

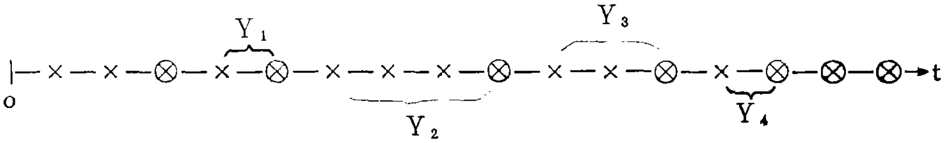
叙述其序列的取法。

（a）**震前序列**　东部地区（第九区）地震前一次第i区地震与下一次东部地区地震的时间间隔。



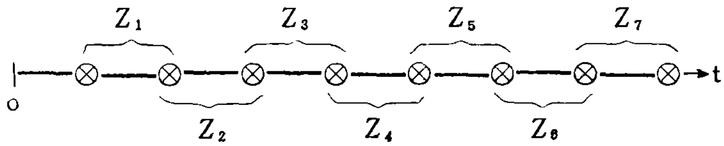
在$\overline{ot}$时间轴上，用⊗标示东部地区的发震时间，用X表示第i区的发震时间，如上所示，则时间间隔序列｛X_n｝，称为第i区的震前序列。

（b）**震后序列**　东部地区地震后一次第i区震与下一次东部地区地震的时间间隔。



$\overline{ot}$时间轴上，⊗，×意义同前，如上所示，则时间间隔序列｛y_n｝称为第i区的震后序列。

（c）**时间间隔序列**　东部地区地震与下一次东部地震的时间间隔。



$\overline{ot}$时间轴上，⊗×意义同前，如上所示，则时间间隔序列｛Z_n｝称为第九区的时间间隔序列。

由于对区域进行了适当的选择，因而这些序列都有其一定的规律，对这些规律作从定性到定量的分析后，认为采取K重（$K\leqslant 3$），正态、平稳自回归模型是比较恰当的，以下将以第四区为例，详细地阐述该模型。

第四区：（15北—25北，110东—145东）+（20北—25北，120东—125东）包括我国台湾省及菲律宾、加罗林群岛等。1900年至1976年，该地区共发生175次震（震级按前面规定）。照（a）构造震前序列｛x_n｝如下（以天为单位）：

425，2230，400，838，413，45，365，189，163，127，218，281，17，440，190，500，294，752，415，500，392，708，687，627，439，214，197，416，587，60，242，1382，217，144，1120，462，669，144，623，276，666，177，155。

将这些数据点在（n，t）轴上，n表示发震次数，t表示x_n的值。

由上图可以看出，它们集中在400周围，並作随机摆动，无明显趋势，因而可以只用随机模型处理它。

一、正态性检验　将这些数据分为5组，作X^2检验，自由度$n=5-3=2$，信度$\alpha=0.01$，$X^2_{0.01}=11.345$，实际计算为10.42，小于11.345，故可认为该序列是正态的。

二、平稳性检验

①检验数学期望EX_n是否近似一常数：将数据分成S组，设第i组的总数为n，计算$\overline{X}_i=$

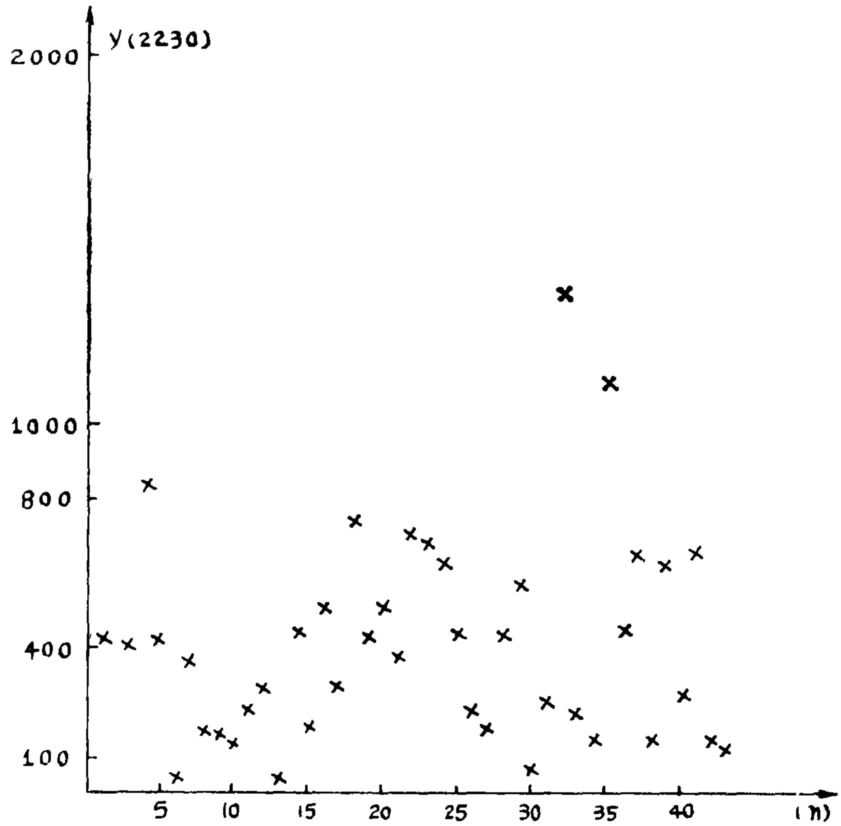


图 一

$\frac{1}{n_i}\sum_{K=1}^{n_i} X_{i+K}$ ($i=1, 2, \cdots\cdots s$) 发现 $\overline{X}_i$ 波动不大，故可近似地认为 EX_n 是一常量。

②$R(n, m)=E(X_n-EX_n)(X_{n+m}-EX_{n+m})$ 是否与n无关。

将数据用截头去尾，拦腰切取，分段等办法，取得若干子组，并利用样本协方差函数：

$$\gamma(n, m)=\frac{1}{N-m}\sum_{K=1}^{N-m}(X_{K+n}-\overline{X}_n)(X_{K+m+n}-\overline{X}_n)$$

其中 $\overline{X}_n=\frac{1}{N}\sum_{K=1}^{N}X_{K+n}$ N—组内样本个数。对每一子组计算其值。发现，各子组样本协方差函数：

$$\gamma(n, m)=\frac{1}{N-m}\sum_{K=1}^{N-m}(X_{K+n}-\overline{X}_n)(X_{K+m+n}-\overline{X}_n)$$

其中 $\overline{X}_n=\frac{1}{N}\sum_{K=1}^{N}X_{K+n}$ N—组内样本个数。对每一子组计算其值，发现各子组样本协方差函数大体相似（当m相同时），故可认为相关函数与n无关。

三、S重相关性检验　由于该序列是平稳、正态的。检验其S重相关性，转化为检验序

列$\{X_n\}$与$\{X_{n+K}\}$当K≤S时相关，当K>S时不相关。

用公式：

$$R_K=\frac{2\sum_i(X_{i-K}-\overline{X}_K)(X_i-\overline{X}_o)}{\sum_i(X_{i-K}-\overline{X}_K)^2+\sum_i(X_i-\overline{X}_o)^2}$$

其中 $\overline{X}_o=\frac{1}{n-K}\sum_{i=K+1}^{n}X_i \quad \overline{X}_K=\frac{1}{n-K}\sum_{i=1}^{n-K}X_i$

进行检验[1]，其结果当k=1，2时，相关（取显著水平$2\alpha=0.01$），K=3，4，5，6不相关，因而可以近似地认为，此序列是二重相关的。

通过上述各种检验，对该序列采用二重、正态、平稳自回归模型是比较合适的。设该序列的期望值$EX_n=a$，相关函数

$$R(n)=E(X_m-a)(X_{m+n}-a)$$

又补设X_n是一个遍历的平稳序列，有数学期望和相关函数的统计估值：

$$a\approx\overline{X}=\frac{1}{N}\sum_{j=1}^{N}X_j$$

$$R(n)\approx Y(n)=\frac{1}{N-n}\sum_{j=1}^{N-n}(X_j-\overline{X})(X_{j+n}-\overline{X})$$

其中 N——样本函数，在上述例中N=43。

代替序列$\{X_n\}$，考虑均值为零的序列$\{X'_n\}=\{X_n-a\}$，$\{X'_n\}$也是二重、正态、平稳自回归模型，因而应有：

$X'_n=a_1X'_{n-1}+a_2X'_{n-2}+Z'_n$，其中$Z'_n$是噪声，令$X'^*_n=a_1X'_{n-1}+a_2X'_{n-2}$，利用正态、平稳过程的理论[2]，系数$\alpha_1$，$\alpha_2$可按下述原则定出：使误差（$X'_n-X^*_n$）平方的期望值$E(X'_n-X'^*_n)^2$达到最小。易证，系数$\alpha_1$、$\alpha_2$应满足2阶线性代数方程组，

$$\begin{pmatrix}R(o) & R(1)\\ R(1) & R(o)\end{pmatrix}\begin{pmatrix}a_1\\ a_2\end{pmatrix}=\begin{pmatrix}R(1)\\ R(2)\end{pmatrix}$$

（实际计算时，R(n)，a，均用估值代替）

由上机计算得，$a_1=0.37$，$a_2=0.078$

故 $X'^*_n=0.37X'_{n-1}+0.078X'_{n-2}$

利用X_n与X'_n关系得：

$$X^*_n-a=0.37(X_{n-1}-a)+0.078(X_{n-2}-a)$$

$$X^*_n=(a-0.37a-0.078a)+0.37X_{n-1}+0.078X_{n-2}$$

以$\overline{X}$代替a，计算得上式右方第一项为249，从而得到：

$$X_n=249+0.37X_{n-1}+0.078X_{n-2}+Z_n$$

$$X^*_n=249+0.37X_{n-1}+0.078X_{n-2} \tag{A}$$

改写 $X_n=X^*_n+Z_n=X^*_n(1+Z_n/X_n)$

通过定性分析，可将Z_n/X^*_n近似地当作正态序列，该正态分布的期望估值为b=0.603，方

差估值$\sigma=0.51$。

查正态分布表，可得：

$$P(|Z_n/X_n^*-h|<K\sigma) \qquad (B)$$

（B）式的值，其中K是一个待定的非负实数，显然，为了提高报准率，K应取得大些，但K大了预报区间变长了，为了协调这两者的矛盾，对不同区域，根据实际σ大小，选择K，使（B）式的值≥80%，对（4）区，因$\sigma=0.51$，我们选取$K=3/2$，对应的有：

$$P\left(|Z_n/X_n^*-b|\leqslant\frac{3}{2}\sigma\right)\geqslant 0.80 \qquad (C)$$

由（C）推得：

$$P(X_n^*(b-3/2\sigma)\leqslant Z_n\leqslant X_n^*(b+3/2\sigma))\geqslant 0.80$$

$$P((1+b-3/2\sigma)X_n^*\leqslant Z_n+X_n^*\leqslant(1+b+3/2\sigma)X_n^*)\geqslant 0.80$$

即 $$P((1+b-3/2\sigma)X_n^*\leqslant X_n\leqslant(1+b+3/2\sigma)X_n^*)\geqslant 0.80$$

利用实际资料，计算出X_{n-1}，X_{n-2}，代入（A）式计算得X_n^*，将X_n^*代入（D）式，计算出X_n的预报区间：

$$X_n:\ ((1+b-3/2\sigma)X_n^*\ \ (1+b+3/2\sigma)X_n^*) \qquad (E)$$

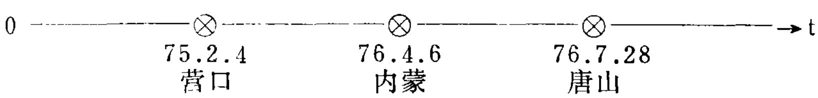
由（D）知，此区间的可靠性不小于0.80。

仿上例作法，求出每一区的“前”，“后”序列的X_n^*的方程，及相应的b，σ，K，从而可计算出预报区间（今后，随着时间的推移，东部发生了新的地震，有关的区域加入新的资料后，再重新计算一次(E)式）。附表

综合模式的形成及预报准则

以下通过几个实例，说明综合预报的问题。

例1 唐山大震



这里，我们需要回答几个问题：①唐山大震的震前、震后序列有哪些？②如何求？③怎样综合这些前（后）序列的预报区间？④前（后）序列在预报中各起什么作用？下面我们将回答这些问题。

首先，找出75年2月4日后，76年4月6日前，上述八个区域内，哪些发生了地震（按规定的震级范围）。发现②，③，④，⑥，⑦，⑧发震了，这些区所对应的震前序列，就是唐山大震的全部震前序列，并按（E）式求出上述各区震前序列的预报区间。

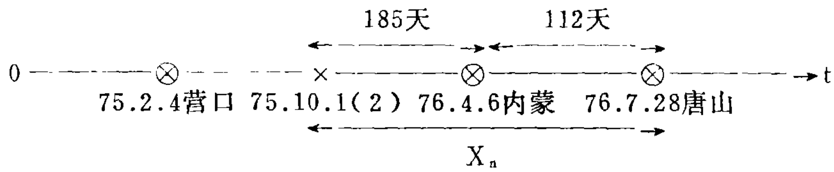
如(2)区： $$X_n=1979-0.82X_{n-1}-0.729X_{n-2}+Z_n$$

$$X_n^*+Z_n=X_n^*(1+Z_n/X_n^*)$$

其相对误差 Z_n/X_n^*的期望估值$b=0.00059$

方差估值$\sigma=0.47$ 取（B）式中$K=3/2$，相应的预报区间：X_n：（95，315）

注意到75年2月4日后，76年4月6日前，②区最后一次震的发震时间是75年10月1日。



为了各区便于比较，把求出的预报区间，都换算成76年4月6日起的间隔时间区间，这只需减去一个时段即可，如（2）区，变为间隔时间τ的区间：（95－185　315－185）以实际意义考虑应取τ：（0　130）

找出其它几个区，在75年2月4日与76年4月6日间，最后一次发震时间，分别为：

③　75、7、20　（6南，155.5东）　7.5

④　76、2、23　（23.4北，121.3东）　6.3

⑥　76、1、21　（47.6北，151.1东）　7

⑦　75、8、15　（54北，168东）　7.2

⑧　76、2、16　（22.4北，100东）　5.8

利用最后一次发震时间，将求出的后震序列的预报区间，都仿（2）换算成间隔时间τ的区间，分别为：

②　（0　130）

③　（0　210）

④　（27　210）

⑥　（15　337）

⑦　（0　315）

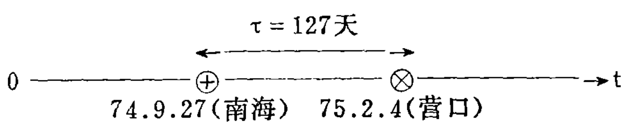
⑧　（103　301）

东部本身的间隔时间区间为⑨（67　337），如何将它们综合起来？在这些震前序列中，找出（B）式中K最大者（即σ最小者），所对应的间隔时间区间作为趋势预报时段，在该例中，取②区，即（0　130）范围为基本时段，以它们共同（或大部分）相交部分，作为第一个重点预报时间，在此时间内，若未发震，再利用后面一个共同部分作为第二个重点预报时间，若这样的第二个共同部分不存在，则利用后震序列补充。在上述例中，第一个重点预报时间是（103，130）实发在此范围内。

随着时间的推移，76年4月6日后，震后序列也不断出现，至7月28日前，共出现①、②、③、⑧，四个区。原则上，应求出这些区的震后序列的预报区间，并转化为时间间隔τ的区间，求共同部分，给出另一个重点预报时间，但在此例中，第一个重点时间不长（不足一个月），而实发在此区间内，因而震后序列也就不必一一求出了。

例2 营口震

仿例1，求出全部前震序列相应的间隔时间范围：



⑥　（100　329）

⑧　（0　236）

⑨ （87 437）

选取最优者⑥区所对应的区间为趋势范围，并以其共同部分（100 236）作为第一个重点预报时间。实发在此区间内，且经计算，后震序列的预报时段与（100 236）无共同部分，不能起缩短区间的作用，因而该例中震后序列实际未起作用。

例3 76年4月6日内蒙震

震前序列对应的时间间隔范围分别是：

③ （0 208）

⑧ （0 503）

⑨ （110 552）

以最优者③作为趋势时段，以共同部分（110 208）作为第一个重点预报时间，该时间段太长，需要研究震后序列的补充、修正。

先将发生在75年2月4日后110天内的震后序列的间隔时间列出：

① （303 575）

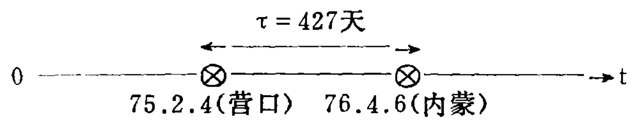
② （277 438）

③ （100 381）

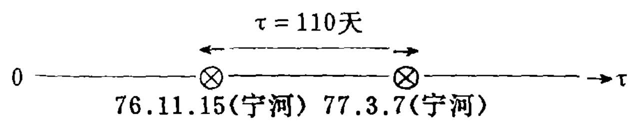
④ （114 572）

⑧ （147 965）

震后序列与震前③的共同部分（147 208）缩短了第一个重点预报时段，因而用它作为修正后的第一个重点预报时间，但实际上在上述时段内未发震，必需考虑补充时间，震后序列的共同部分：（303 438）以此作为第二个重点预报时间，实发在此时段内。

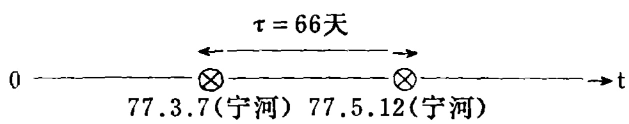


77年3月7号震后验情况：



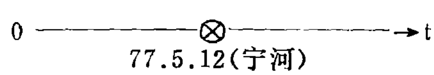
予报区间τ：（87 207）实发在此区间。

77年5月12号震的后验情况



予报τ：（62 170）实发在此区间内。

77年5月12号，实报情况



预报τ：（183 157）地点30北以北（从后面地点计算可知）实发：1977.11.27，MS = 5.7，τ = 195，地点宁河，与预报相符。

地点预报公式

将东部地区以30°北为界，划分为南、北两部分，并使其0，1化，北为1，南为0，这样使1900年—1976年的东部地震，构成了一个0，1序列记为$\{Z_n\}$。

从前面时间预报的规则中可以看到，对应于每一次东部地震，九个区域的震前（后）序列不一定都出现，对应于第n次东部震，若第i区震前序列出现，则记为1，否则记为0，为区别起见，令

$$X_i(n)=\begin{cases}1 & \text{出现}\\ 0 & \text{否}\end{cases}$$

8个前震序列，按此方式0，1化，得到预报地点的8个因子，利用概率回归估计法，得到预报地点的"前"预报方程：

$$Z_1=0.627+0.00348Z_1+0.061X_2+0.0402X_3-0.217X_4-0.0841X_5+0.0926X_6+0.136X_7+0.00615X_8$$

其中，X_1—（4）区　X_2—（3）区　X_3—（7）区

X_4—（5）区　X_5—（6）区　X_6—（1）区

X_7—（2）区　X_8—（8）区

仿上，将震后序列也0、1化得"后"预报方程

$$ZZ=0.629+0.0142X_1+0.116X_2-0.141X_3-0.232X_4-0.0819X_5-0.0418X_6+0.164X_7+0.327X_8$$

其中　X_1—（1）区　X_2—（2）区　X_3—（8）区　X_4—（4）区

X_5—（5）区　X_6—（6）区　X_7—（3）区　X_8—（7）区

根据历史符合情况，定出两个临界值，"前"为$P_1=0.65$当计算出的Z值大于或等于P_1时，报1，否则报0。"后"为$P_2=0.62$，大于或等于P_2时报1，否则报0。

预报规则

1.以"前"预报方程为主。

2.当"前"预报方程的值，在0.5与0.7之间时，参考"后"预报方程的值，若后预报方程的值小于0.4则报0，大于0.8则报1。

例：1976、7、28唐山大震

"前"预报方程　$Z=0.00348+0.061+0.0402-0.0841+0.627+0.136+0.00615=0.79$

根据上述准则应报1，实发1。

1977年5月12日后，下次东部震发震地点预报：前预报方程

$$Z=0.627-0.0841+0.00348+0.0926+0.062=0.705$$

后预报方程　$ZZ=0.629-0.232+0.164+0.327-0.141=0.75$

应报1，实际上，77年11月27日在宁河发生$M_S=5.7$级震，与预报符合。内符效果80%。

附（一）

一区 $\begin{pmatrix}20北—50北\\12东—73东\end{pmatrix}$ 二区 $\begin{pmatrix}26北—10南\\85东—110东\end{pmatrix}$ 三区 $\begin{pmatrix}15北—20南\\110东—160东\end{pmatrix}$

四区 $\begin{pmatrix}20北—25北\\120东—125东\end{pmatrix}+\begin{pmatrix}15北—25北\\110东—145东\end{pmatrix}$

五区 $\begin{pmatrix}25北—40北\\130东—160东\end{pmatrix}$ 六区 $\begin{pmatrix}40北—60北\\140东—160东\end{pmatrix}$ 七区 $\begin{pmatrix}45北—60北\\160东—140西\end{pmatrix}$

八区 我国西部 九区 我国东部

附（二） 各区“震前”、震后序列的自相关函数及相应的 b、σ 值（K一般取为3/2）或误差范围*。

一区

“震前” $X_n^* = 431 + 0.23X_{n-1} + 0.066X_{n-2}$

$b = 0.03$ $\sigma = 0.8$ 〔−0.8, 0.8〕 81%

“震后” $X_n^* = 502 + 0.042X_{n-1} - 0.11X_{n-2}$

〔−0.8, 0.8〕 79%

二区

“震前” $X_n^* = 1979 - 0.82X_{n-1} - 0.729X_{n-2}$

$b = 0.00059$ $\sigma = 0.47$ 〔−0.8, 0.8〕 94%

“震后” $X_n^* = 443 - 0.26X_{n-1} - 0.14X_{n-2}$

〔−0.8, 0.8〕 77%

三区

“震前” $X_n^* = 181 + 0.374X_{n-1} + 0.159X_{n-2}$

$b = 0.619$ $\sigma = 0.81$ 〔−0.8, 0.8〕 82%

“震后” $X_n^* = 333 + 0.186X_{n-1} + 0.001X_{n-2}$

$b = 0.767$ $\sigma = 0.62$ 〔−0.8, 0.8〕 83%

四区

“震前” $X_n^* = 249 + 0.367X_{n-1} + 0.078X_{n-2}$

$b = 0.603$ $\sigma = 0.51$

“震后” $X_n^* = 540 - 0.018X_{n-1} + 0.042X_{n-2}$

$b = 0.956$ $\sigma = 0.601$

五区

“震前” $X_n^* = 295 + 0.432X_{n-1} + 0.014X_{n-2}$

$b = 0.681$ $\sigma = 0.73$

“震后” $X_n^* = 241 + 0.139X_{n-1} + 0.211X_{n-2}$

$b = 0.590$ $\sigma = 0.46$

六区

“震前” $X_n^* = 331 + 0.429X_{n-1} - 0.085X_{n-2}$

〔−0.8, 0.8〕 84%

“震后” $X_n^* = 401 + 0.172X_{n-1} - 0.0194X_{2}$

〔−0.8, 0.8〕 80%

七区

"震前" $X_n^* = 324 + 0.169X_{n-1} + 0.175X_{n-2}$

〔-0.8, 0.8〕 78%

"震后" $X_n^* = 362 + 0.0932X_{n-1} + 0.0509X_{n-2}$

〔-0.8, 0.8〕 77%

八区

震前 $X_n^* = 451 + 0.082X_{n-1} + 0.04X_{n-2}$

〔-0.8, 0.8〕 81%

震后 $X_n^* = 425 + 0.12X_{n-1} + 0.06X_{n-2}$

〔-0.8, 0.8〕 78%

九区

$X_n^* = 244 + 0.11X_{n-1} + 0.16X_{n-2}$

〔-0.8, 0.8〕 76%

（1980年7月收到）

参 考 资 料

〔1〕地震发生时间的概率预报（一） 《地球物理学报》 17卷1期

〔2〕王梓坤著 随机过程论 第九章 第三节

〔本文得到王梓坤教授的指导和帮助，在此表示衷心的感谢。〕

* 以一区"震前"序列为例，所谓误差范围是〔$-0.8$0.8〕—81%即指：$P(-0.8 \leq X_n/X_n^* \leq 0.8) \approx 81\%$ 因而它相当于给出了b，σ和K。