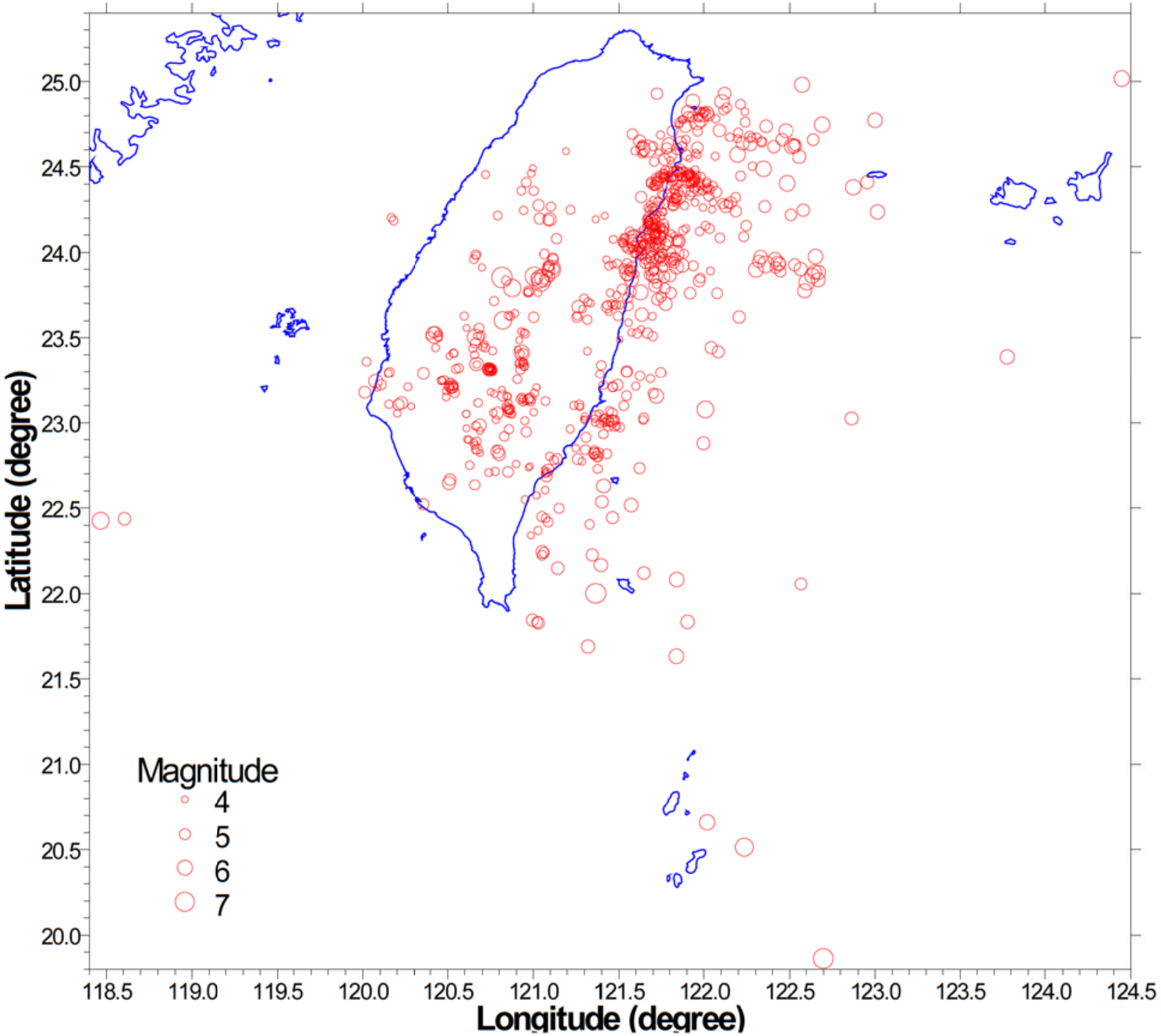
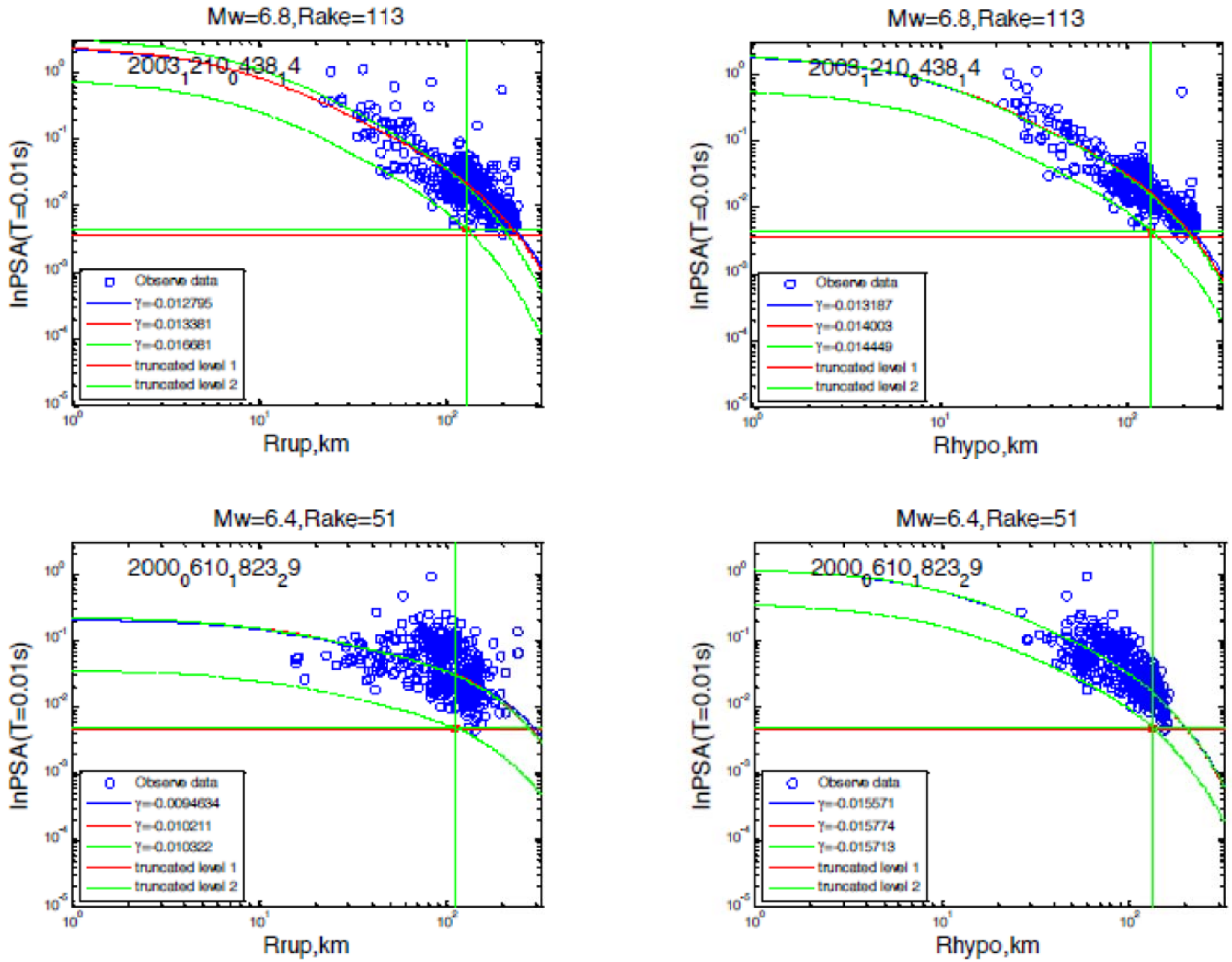
地震風險評估的統計模型與分析

程毅豪 中研院統計科學所
蔡主權 中研院地球科學所
趙書賢 國家地震工程研究中心

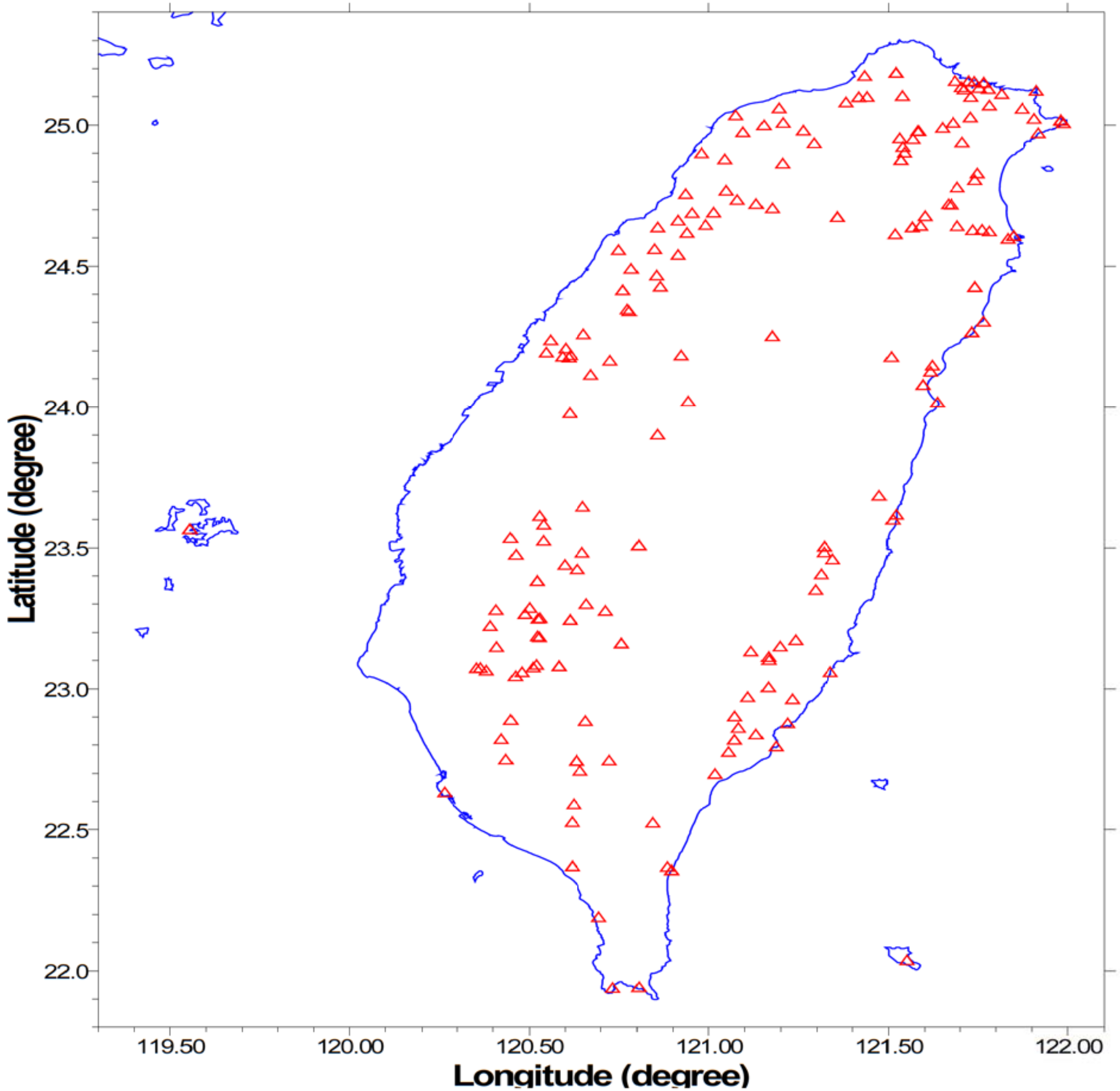
Earthquake: Major Threat to Us



Ground Motion Attenuation



Data Collection: Strong Motion Stations



Peak Ground Acceleration (PGA) attenuation model

$$y_i = g(M_i, R_i; \theta) + \varepsilon_{ei} + \varepsilon_{si} + \varepsilon_{ri}$$

- y_i : $\log_{10}$ PGA
- M_i : local magnitude
- R_i : hypocentral distance (source-to-site distance)
- θ : regression parameter
- g : known scalar function

- ε_{ei} : event-specific term, representing the earthquake-to-earthquake variation
- ε_{si} : site-specific term, representing the site-to-site variation
- ε_{ri} : record-specific term, representing the residual variation

$\beta_e = (\varepsilon_e^{(1)}, \cdots, \varepsilon_e^{(E)})'$: collection of E values of ε_e

$$\varepsilon_{ei} = z_i \beta_e$$

$$z_i = (z_{1i}, \cdots, z_{Ei})$$

where

$$z_{ji} = \begin{cases} 1, & \text{if record } i \text{ comes from earthquake } j \\ 0, & \text{otherwise} \end{cases}$$

$$(j = 1, \cdots, E)$$

$\beta_s = (\varepsilon_s^{(1)}, \cdots, \varepsilon_s^{(S)})'$: collection of S values of ε_s

$$\varepsilon_{si} = w_i \beta_s$$

$$w_i = (w_{1i}, \cdots, w_{Si})$$

where

$$w_{li} = \begin{cases} 1 \ , & \text{if record } i \text{ comes from site } l \\ 0 \ , & \text{otherwise} \end{cases}$$

$$(l = 1, \cdots, S)$$

Random Effects

$$\beta_e \sim \mathcal{N}_E(0, \sigma_e^2 I_E)$$

$$\beta_s \sim \mathcal{N}_S(0, \sigma_s^2 I_S)$$

$$\overset{iid}{\varepsilon_{ri} \sim \mathcal{N}(0, \sigma_r^2)}$$

Alternative representation of the model — Mixed effect model

$$y_i = g(M_i, R_i; \theta) + z_i \beta_e + w_i \beta_s + \varepsilon_{ri}$$

$$\equiv G(x_i; \beta) + \varepsilon_{ri}$$

$$x_i = (M_i, R_i, z_i, w_i)$$

$$\beta = (\theta', \beta_e', \beta_s')'$$

Estimation Procedure & Computation Algorithm

- MLR estimation (Dempster et al., 1981 JASA)--

  treating fixed effects as random, but with infinite covariance matrix, and then finding MLE for the variance components $\eta = (\sigma_r^2, \sigma_e^2, \sigma_s^2)$

- EM algorithm (Dempster et al., 1977)--

  treating (Y, β) as the 'complete' (augmented) data, the MLE of η based on the actual data Y can be obtained via EM.

- Dempster et al., (1981 JASA) have proposed EM for MLR estimates in linear variance components models

- For linear random effects model:

$$\widetilde{Y}_{N\times 1} = \widetilde{X}_{N\times M}\beta_{M\times 1} + \varepsilon_{N\times 1}$$

$$\beta \sim \mathcal{N}_M(0,\Sigma)$$

$$\varepsilon \sim \mathcal{N}_N(0,\sigma^2 I_N)$$

$$\beta \perp \varepsilon$$

$$\Sigma = \mathrm{diag}\{\sigma_1^2 I_{n_1},,,\sigma_K^2 I_{n_K}\}$$

$$\sum_{k=1}^{K} n_k = M$$

$$\Rightarrow$$

$$\beta \mid \widetilde{Y} \sim \mathcal{N}_M(\hat{\beta},\Omega)$$

$$\hat{\beta} = P^{-1}\widetilde{X}'\widetilde{Y}$$

$$\Omega = \sigma^2 P^{-1}$$

$$P = \widetilde{X}'\widetilde{X} + \sigma^2\Sigma^{-1}$$

- E-Step--
  compute the conditional expectations of complete-data sufficient statistics given actual data at current parameter values :

$$E(\beta_k'\beta_k \mid \widetilde{Y};\Sigma,\sigma^2) = \hat{\beta}_k'\hat{\beta}_k + tr(\Omega_{kk})$$

$$E(\varepsilon'\varepsilon \mid \widetilde{Y};\Sigma,\sigma^2) = \widetilde{Y}'\widetilde{Y} - 2\widetilde{Y}'\widetilde{X}\hat{\beta} + \widetilde{\beta}'\widetilde{X}'\widetilde{X}\hat{\beta} + tr(\widetilde{X}'\widetilde{X}\Omega)$$

- M-Step--
  update parameter values by obtaining the complete-data MLEs with the involved sufficient statistics replacing by their conditional expectations :

$$\hat{\sigma}^2 = \frac{E(\varepsilon'\varepsilon \mid \widetilde{Y}, \Sigma, \sigma^2)}{N}$$

$$\hat{\sigma}_k^2 = \frac{E(\beta_k' \beta_k \mid \widetilde{Y}, \Sigma, \sigma^2)}{n_k}$$

- Iterate between E- and M- steps until convergence.

Proposed Algorithm for General Mixed (Variance Components) Models of PGA Attenuation Relationship

$$y_i = G(x_i;\beta) + \varepsilon_{ri}; \qquad Y = (y_1, \cdots, y_N)'$$

$$G(X;\beta) = \{G(x_1;\beta), \cdots, G(x_N;\beta)\}'$$

$$\beta = (\theta', \beta_e', \beta_s')' \sim \mathcal{N}_M(0, \Sigma), \quad M = F + E + S$$

$$\Sigma = \text{diag}(\Sigma_{\theta\theta}, \sigma_e^2 I_E, \sigma_s^2 I_S)$$

where $\Sigma_{\theta\theta}$ is infinite so that

$$\Sigma^{-1} = \text{diag}(O_F, \sigma_e^{-2} I_E, \sigma_s^{-2} I_S)$$

$O_F = F \times F$ matrix of 0's

1. Start with initial values for $\eta = (\sigma_r^2, \sigma_s^2, \sigma_e^2)$ and β; call them current parameter values.
2. Generate the working design matrix

$$\widetilde{X} = \frac{\partial G(X;\beta)}{\partial \beta}$$

and the working outcome vector

$$\widetilde{Y} = \widetilde{X}\beta + \{Y - G(X;\beta)\};$$

both are evaluated at current parameter values.

3. Compuete conditional expectation and covariance of β at current parameter values :

$$\hat{\beta} = P^{-1}\widetilde{X}'\widetilde{Y}$$

$$\Omega = \sigma_r^2 P^{-1}$$

where

$$P = \tilde{X}'\tilde{X} + \sigma_r^2 \Sigma^{-1}$$

$$\Sigma^{-1} = \mathrm{diag}\{O_F, \sigma_e^{-2} I_E, \sigma_s^{-2} I_S\}$$

call $\hat{\beta}$ the current value of β

4. Get the updated value of $\eta = (\sigma_r^2, \sigma_s^2, \sigma_e^2)$ by

$$\sigma_r^2 = \frac{\{\tilde{Y}'\tilde{Y} - 2\tilde{Y}'\tilde{X}\hat{\beta} + \hat{\beta}'\tilde{X}'\tilde{X}\hat{\beta} + tr(\tilde{X}'\tilde{X}\Omega)\}}{N}$$

$$\sigma_e^2 = \frac{\{\hat{\beta}_e'\hat{\beta}_e + tr(\Omega_{ee})\}}{E}$$

$$\sigma_s^2 = \frac{\{\hat{\beta}_s'\hat{\beta}_s + tr(\Omega_{ss})\}}{S}$$

where $\hat{\beta}_e$ and $\hat{\beta}_s$ are conditional means of β_e and β_s, obtained directly from subvertors of $\hat{\beta}$, and Ω_{ee} and Ω_{ss} are corresponding submatrices of Ω

5. Call the updated value of η the current value of η.
Repeat steps 2 - 4 until η attains convergence.

$$\left(\left| \frac{\text{New}\,\eta - \text{Old}\,\eta}{\text{Old}\,\eta} \right| \leq 10^{-5} \right)$$

When reaching convergence,

- variance components estimates $\hat{\eta} = (\hat{\sigma}_r^2, \hat{\sigma}_e^2, \hat{\sigma}_s^2)$
- $\hat{\theta}$, obtained from subvector of $\hat{\beta} = (\hat{\theta}', \hat{\beta}_{\mathrm{e}}', \hat{\beta}_{\mathrm{s}}')'$, gives the MLE of θ; its covariance is given by $\Omega_{\theta\theta}$
- predicted value of Y, $\log_{10}$ PGA, at given M and R:

$$\hat{Y} = g(M, R; \hat{\theta})$$

- 95% C.I. of $\hat{Y}$:

$$\hat{Y} \pm 1.96(\mathrm{var}\,\hat{Y})^{1/2}$$

$$\mathrm{var}\,\hat{Y} = \tilde{X}_\theta \Omega_{\theta\theta} \tilde{X}_\theta' + \hat{\sigma}_r^2 + \hat{\sigma}_e^2 + \hat{\sigma}_s^2$$

$$\tilde{X}_\theta = \text{the first } F \text{ elements of } \tilde{X}$$

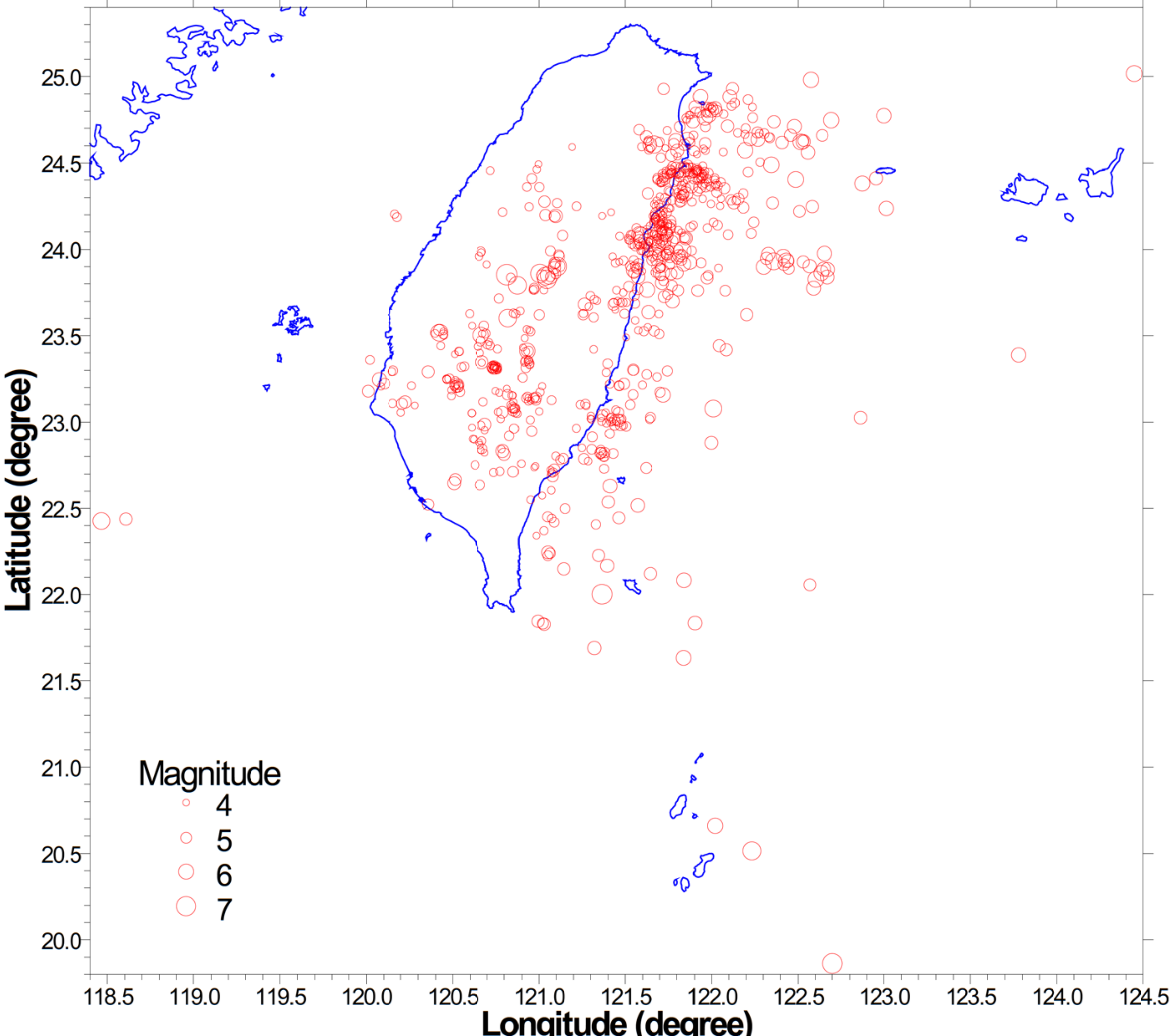
- explicit formula for the covariance (standard errors) of $\hat{\eta}$ (variance components estimates)

  can be obtained (Chen and Tsai 2002 BSSA)

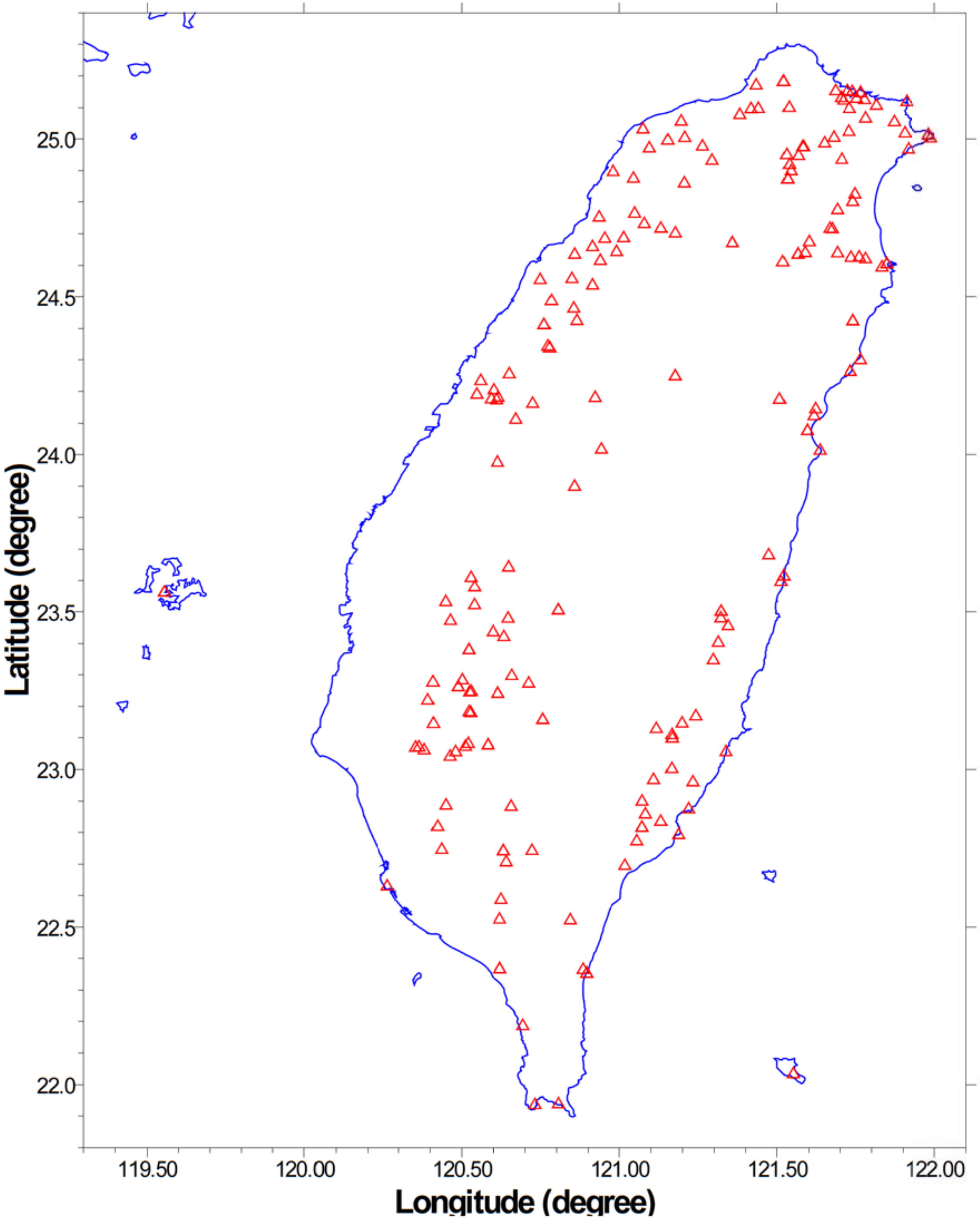
Taiwan Seismic Data

- Database consists of 7123 PGA records from 204 rock-site stations operated by CWB and IES for 744 earthquakes in the Taiwan region during 1976-2000.

  (only sites with more than one event and events with more than one record are included)



Latitude (degree)
Longitude (degree)
25.0
24.5
24.0
23.5
23.0
22.5
22.0
21.5
21.0
20.5
20.0
118.5
119.0
119.5
120.0
120.5
121.0
121.5
122.0
122.5
123.0
123.5
124.0
124.5
Magnitude
4
5
6
7



25.0
24.5
24.0
23.5
23.0
22.5
22.0
Latitude (degree)
119.50
120.00
120.50
121.00
121.50
122.00
Longitude (degree)

The Regression Model

$$g(M,R;\theta) =$$

$$\theta_0 + \theta_1 M + \theta_2 M^2 + \theta_3 R + \theta_4 \log_{10}(R + \theta_5 10^{\theta_6 M})$$

(Fukushima & Tanaka, 1990
Bull.Seism.Soc.Am.; Sharma, 1998
Bull.Seism.Soc.Am.)

Analysis Results

Estimates of Fixed Effects:

θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
0.0004	0.4067	0.7936	-0.0215	-1.7059	5.7848	-0.0565
0.0003	0.4163	0.1622	0.0153	0.8975	5.1533	0.0923

Analysis Results

Estimates of Variance Components:

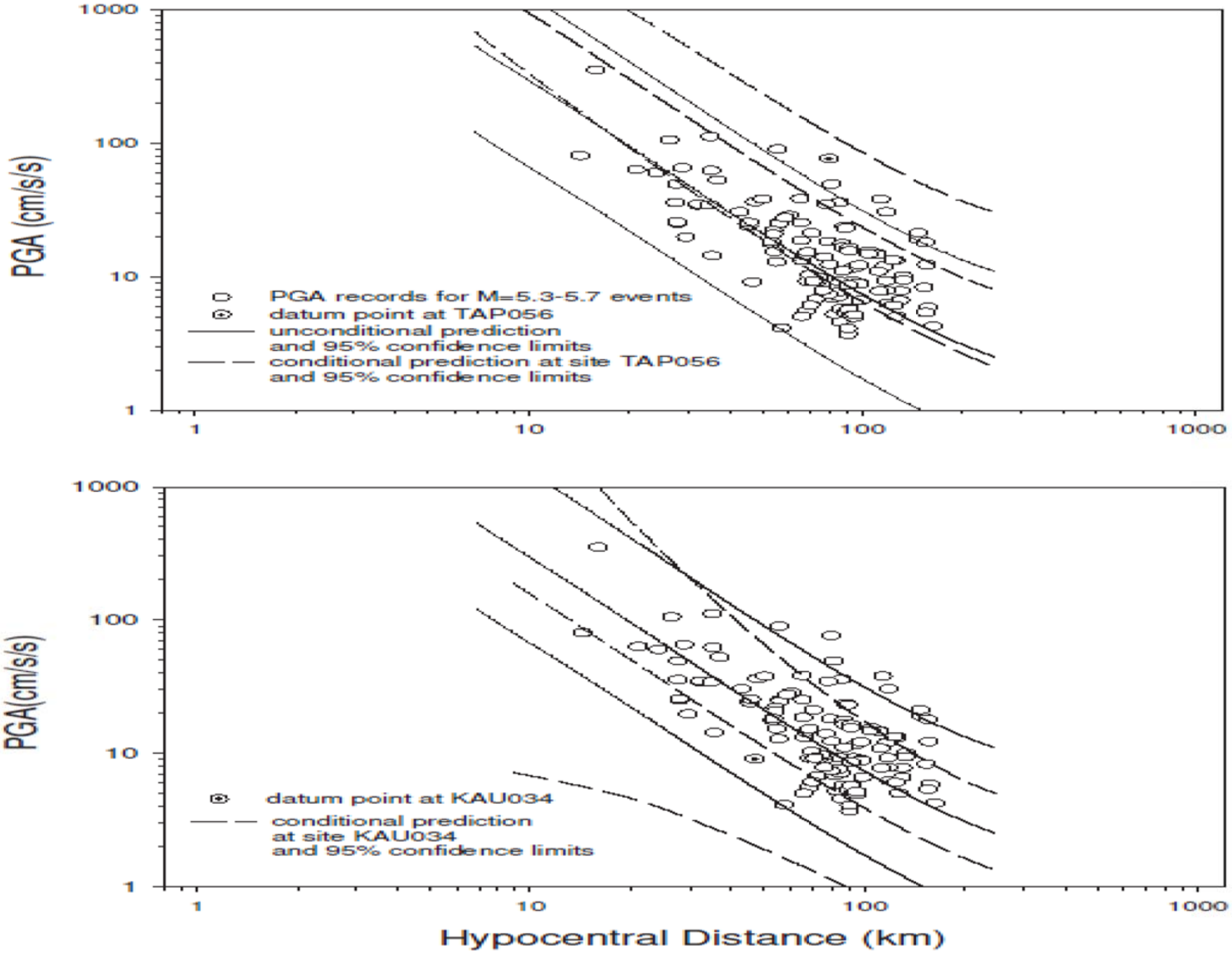
σ_e^2 eq-eq variation	σ_s^2 site-site variation	σ_r^2 residual variation
0.0308	0.0291	0.0397
0.006	0.010	0.002

Predicted Values of β_s

site	β_s
ILA057	-0.60735
HWA026	-0.029552
TAP034	0.056854
KAU049	0.081148
ILA050	0.430644

Predicted Values of β_e

1994/07/11 (M=4.0) (121.8548, 24.4392)	-0.54727
1999/09/20 (M=7.3) (120.8155, 23.8525)	0.056999
1986/11/14 (M=5.3) (121.8120, 23.9630)	0.408006
1993/07/25 (M=4.53) (122.1052, 24.8825)	0.411229



1000
100
10
1
1
10
100
1000
PGA (cm/s/s)
PGA records for M=5.3-5.7 events
datum point at TAP056
unconditional prediction
and 95% confidence limits
conditional prediction at site TAP056
and 95% confidence limits
PGA(cm/s/s)
datum point at KAU034
conditional prediction
at site KAU034
and 95% confidence limits
Hypocentral Distance (km)

Proposed Algorithm Allowing for Covariate-Dependent Inter-Event Variance Components

$$y_i = G(x_i;\beta) + \varepsilon_{ri} \; ; \qquad Y = (y_1, \cdots, y_N)'$$

$$G(X;\beta) = \{G(x_1;\beta), \cdots, G(x_N;\beta)\}'$$

$$\beta = (\theta', \beta_e', \beta_s')' \sim \mathcal{N}_M(0,\Sigma), \quad M = F + E + S$$

$$\Sigma = \mathrm{diag}(\Sigma_{\theta\theta}, \Sigma_{ee}, \sigma_s^2 I_S)$$

where $\Sigma_{\theta\theta}$ is infinite so that

$$\Sigma^{-1} = \mathrm{diag}(O_F, \Sigma_{ee}^{-1}, \sigma_s^{-2} I_S)$$

$$O_F = F \times F \text{ matrix of 0's}$$

General Inter-Event Variance Components

- Now, for a set of covariates v , we model

$$\Sigma_{ee} = \text{diag}\left(f(v_1;\gamma),\cdots, f(v_E;\gamma)\right)$$

f is some specified function and γ is the associated set of parameters (coefficients)

E is the number of events

Modified Algorithm

- Steps 1-3 (see pages 13-14) are unchanged, except that now in step 3,

$$\Sigma^{-1} = \text{diag}(O_F, \Sigma_{ee}^{-1}, \sigma_S^{-2} I_S)$$

- In Step 4,

$$\sigma_r^2 = \frac{\{\tilde{Y}'\tilde{Y} - 2\tilde{Y}'\tilde{X}\hat{\beta} + \hat{\beta}'\tilde{X}'\tilde{X}\hat{\beta} + tr(\tilde{X}'\tilde{X}\Omega)\}}{N}$$

$$\sigma_s^2 = \frac{\{\hat{\beta}_s'\hat{\beta}_s + tr(\Omega_{ss})\}}{S}$$

$$\gamma = \left(\tilde{V}'W\tilde{V}\right)^{-1}\tilde{V}'WB_e$$

$$\tilde{V} = \frac{\partial F(\gamma)}{\partial \gamma}, F(\gamma) = \{f(v_1;\gamma), \cdots, f(v_E;\gamma)\}'$$

$$W = \mathrm{diag}\left(\frac{1}{f(v_1;\gamma)^2}, \cdots, \frac{1}{f(v_E;\gamma)^2}\right)$$

$$B_e = \tilde{V}\gamma + \left(\hat{\beta}_e^2 + \mathrm{vecdiag}(\Omega_{ee}) - F(\gamma)\right), \hat{\beta}_e^2 = \left(\hat{\beta}_{e1}^2, \cdots, \hat{\beta}_{eE}^2\right)'$$

$\mathrm{vecdiag}(\Omega_{ee})$ *is the vector of diagonal elements of* Ω_{ee}

- Step 5 is the same as before (page 17); that is, the algorithm is iterated between Steps 2-4 until convergence

Further Decomposition

- Extract the path-to-path component from the residual error:

$$y_i = g(x_i;\beta) + \varepsilon_{ei} + \varepsilon_{si} + \varepsilon_{ri}$$

$$= g(x_i;\beta) + \varepsilon_{ei} + \varepsilon_{si} + \varepsilon_{pi} + \varepsilon_{ri}^{*}$$

$$= G(x_i;\beta) + \varepsilon_{pi} + \varepsilon_{ri}^{*}$$

$\varepsilon_{pi} \sim N(0,\sigma_p^2)$: path-to-path error

$\varepsilon_{ri}^{*} \sim N(0,\sigma_r^{*2})$: refined residual error

Bivariate Normal Formula

$$(\varepsilon_p, Y) \sim \text{BivariateNormal}$$

$$\text{Cov}(\varepsilon_p, Y) = \text{Cov}(\varepsilon_p, \varepsilon_p + \varepsilon_r^*) = \sigma_p^2$$

$$E(\varepsilon_p \mid Y) = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_r^{*2}}[Y - G(X;\beta)]$$

$$= \frac{\sigma_p^2}{\sigma_p^2 + \sigma_r^{*2}}(\varepsilon_p + \varepsilon_r^*) = \frac{\sigma_p^2}{\sigma_r^2}\varepsilon_r$$

- We see that

$$E(\varepsilon_p \mid Y) + E(\varepsilon_r^* \mid Y) = E(\varepsilon_r \mid Y) = \varepsilon_r$$

so we can use

$$E(\varepsilon_p \mid Y) = \frac{\sigma_p^2}{\sigma_r^2}\varepsilon_r$$

to estimate the path-to-path error component

- Information on σ_p^2 is required

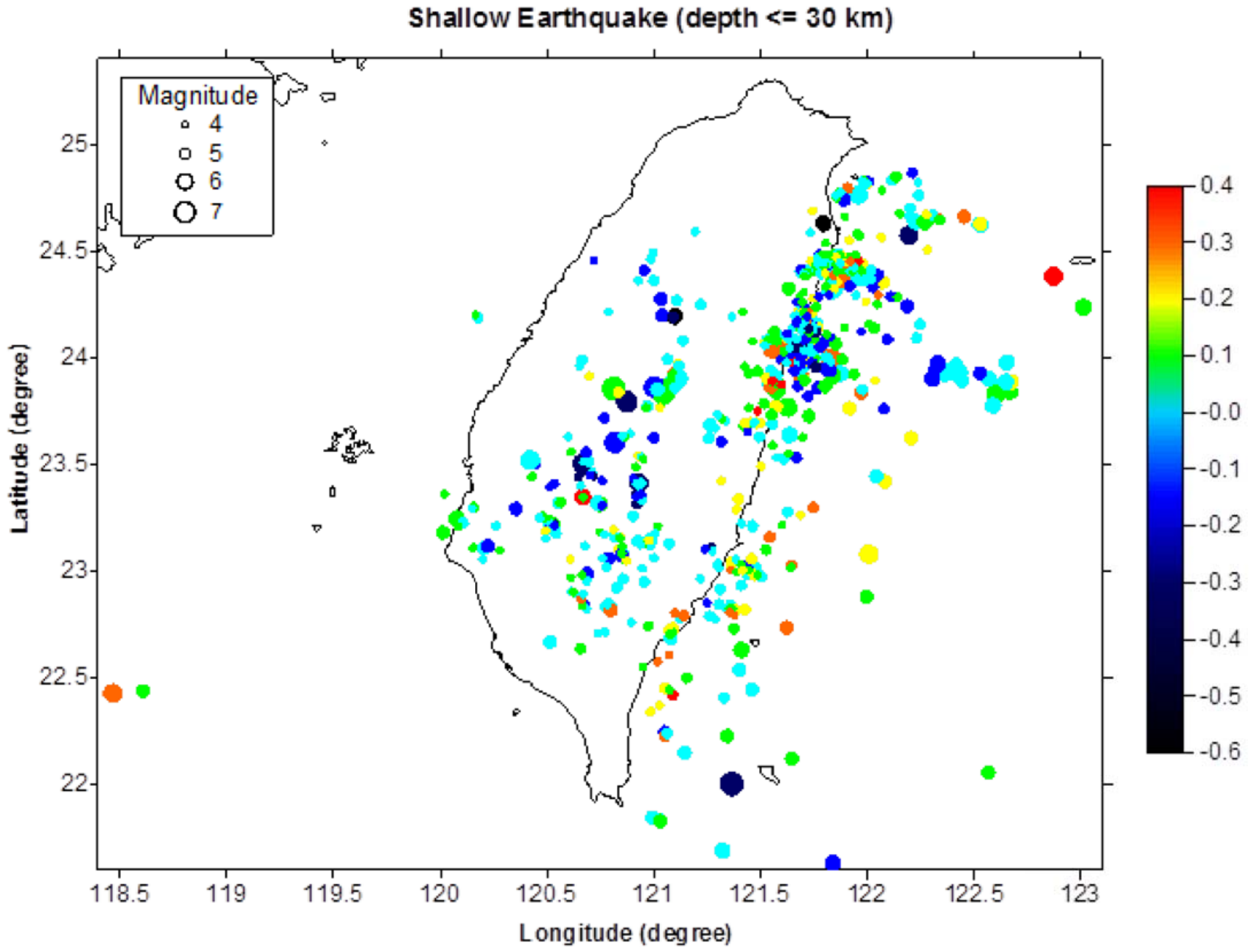
Taiwan PGA Data

- 7123 records collected from 744 earthquake events and 204 sites in Taiwan
- Ground motion model :

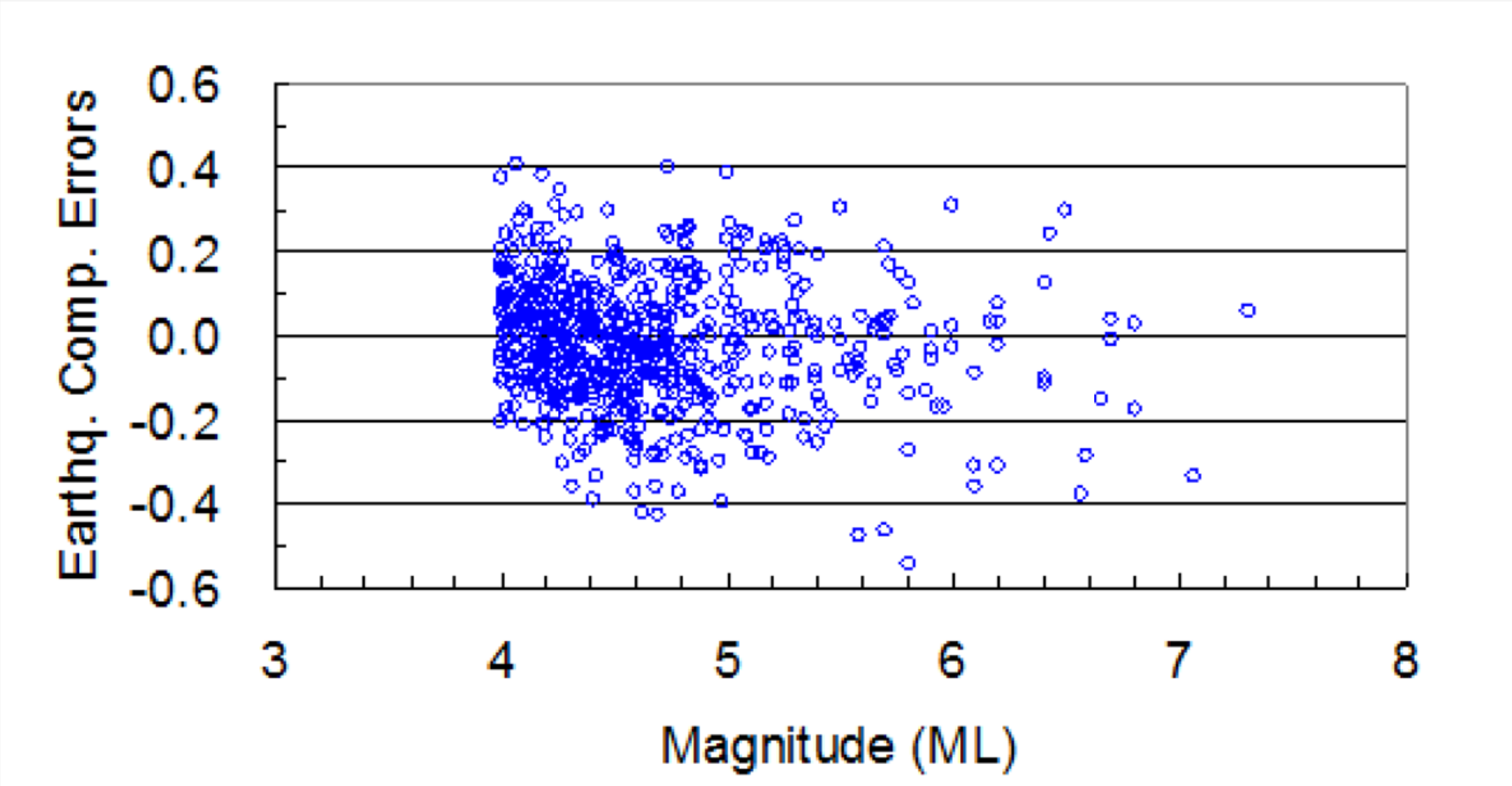
$$\log_{10} PGA = \theta_0 + \theta_1 M + \theta_2 M^2 + \theta_3 R + \theta_4 \log_{10}(R + \theta_5 10^{\theta_6 M})$$
$$+ \varepsilon_e + \varepsilon_s + \varepsilon_r^*$$

- Predicted value of event-to-event error

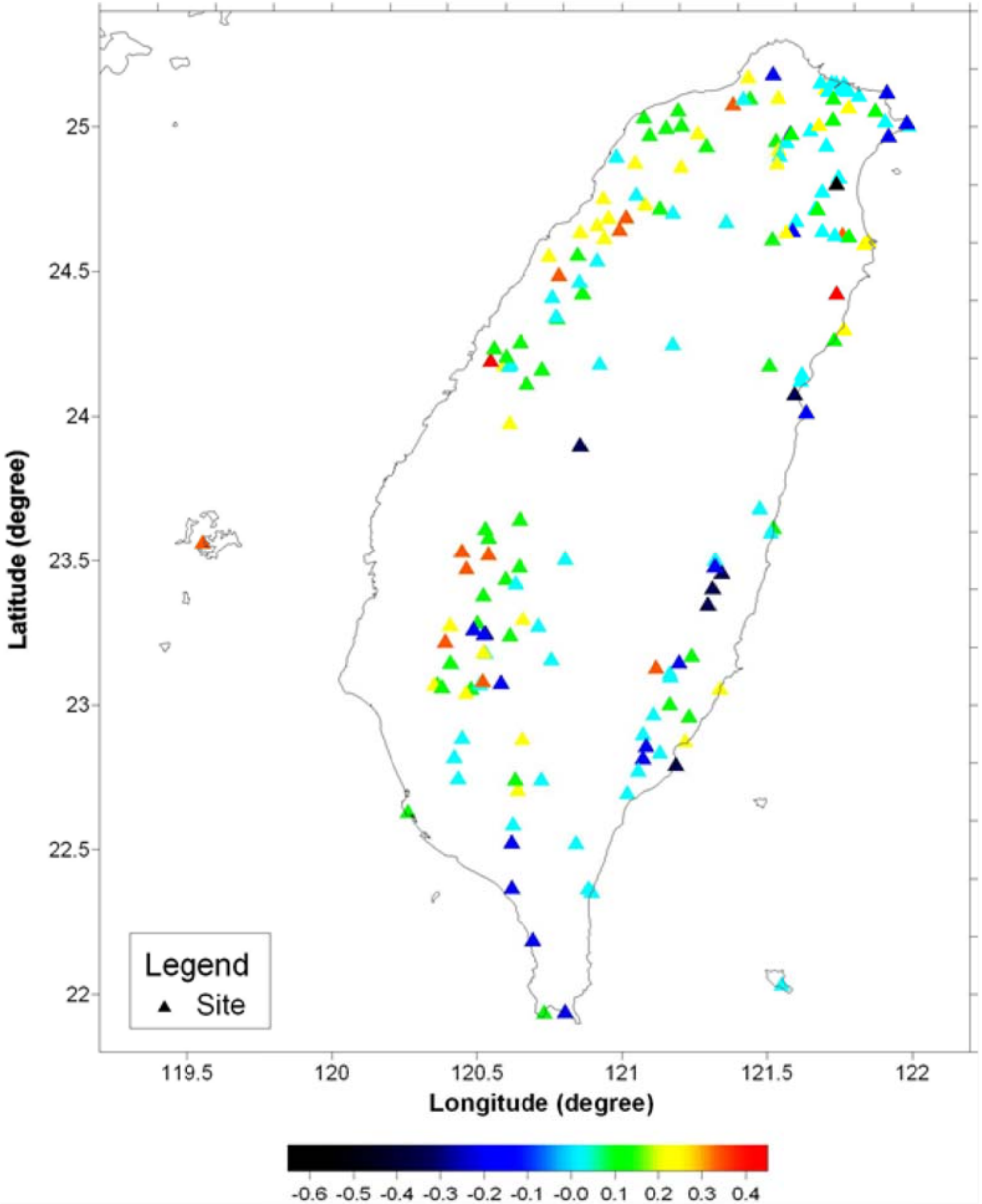
$$\hat{\varepsilon}_e = \hat{E}(\varepsilon_e \mid Y)$$



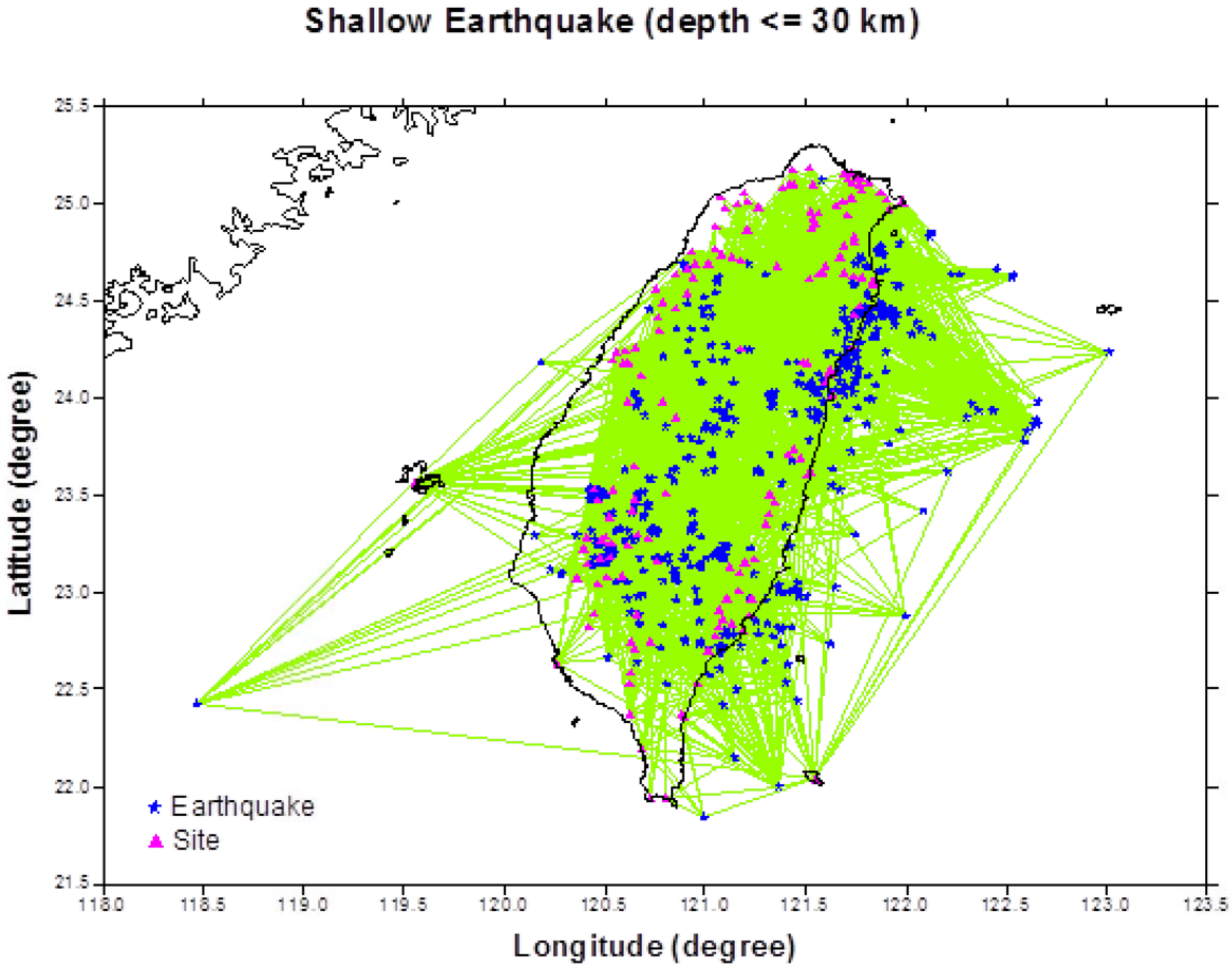
- Earthquake-to-earthquake component of error seems to be unrelated to magnitude (*M*)



- Predicted value of site-to-site error $\hat{\varepsilon}_s = \hat{E}(\varepsilon_s \mid Y)$

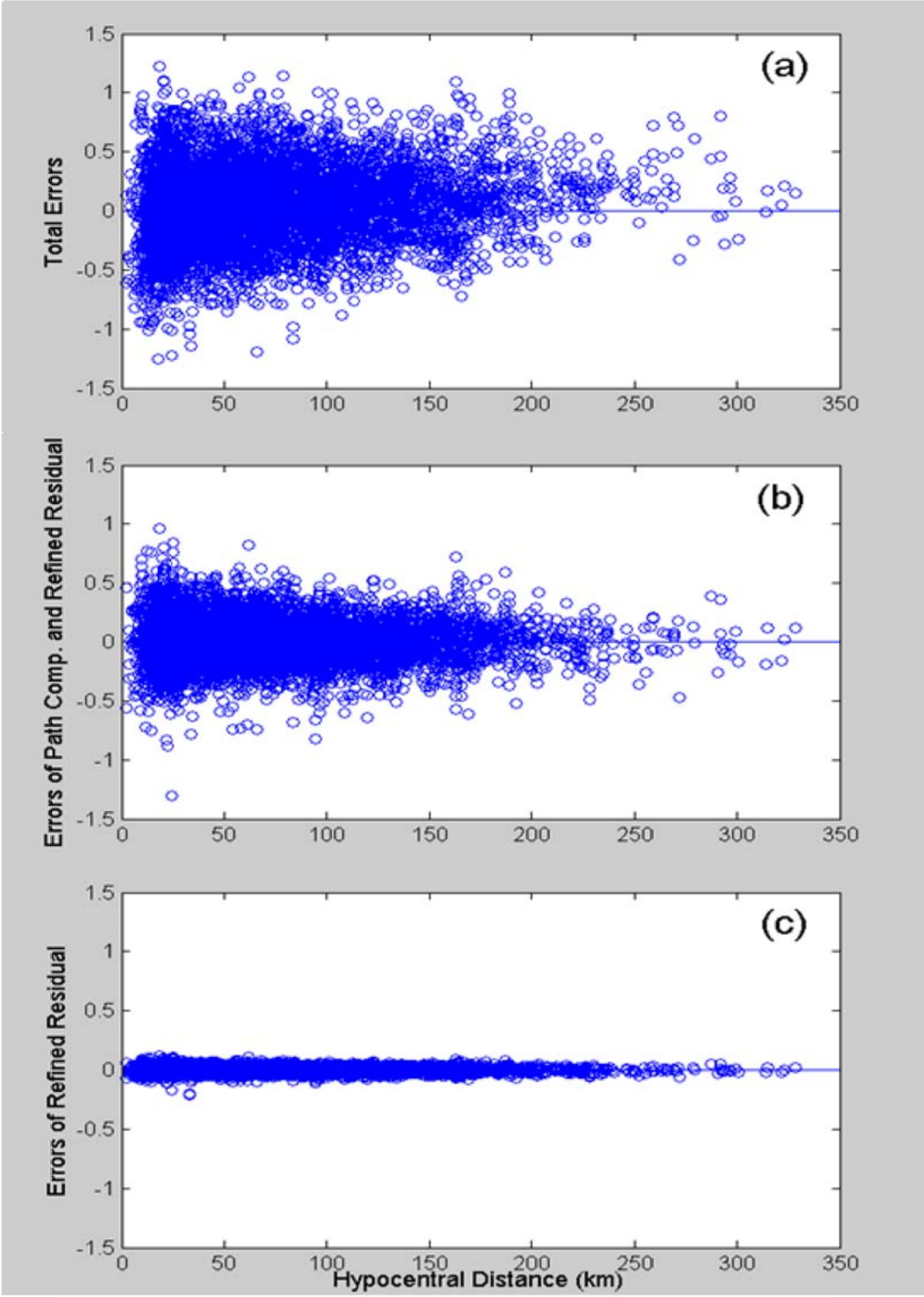


- Paths of earthquake-to-site pairs



- Choose the path-to-path variance component σ_p^2 about the same magnitude of σ_e^2 and σ_s^2 ($\sigma_p^2 \approx 0.03$)
- The refined residual variance component is

$$\sigma_r^{*2} \approx 0.0097$$


(a)
Total Errors
(b)
Errors of Path Comp. and Refined Residual
(c)
Errors of Refined Residual
Hypocentral Distance (km)

Conclusions

- Since the 1999 Chi-Chi earthquake, strong-motion recordings in Taiwan have increased tremendously
- Database for Taiwan strong-motion records is available

- Statistical regression models have been widely applied to prediction of ground motion attenuation for risk assessments of seismic hazards

- More complex models, such as mixed (random effects) models can be applied to repeated measurements of ground motion data for dissection of various sources of ground motion variability
- More accurate prediction can be achieved by accounting for known sources of variance components

- Statistical algorithms such as the EM algorithm can be applied to analysis of ground motion attenuation models for stable and efficient computation
- The EM algorithm together with the MLR procedure is especially convenient for generalization to different ground motion attenuation models with different random/fixed effects and models for the variance components

- Spatial correlation may be further incorporated into the variance-covariance matrix of the random effects
- Statistical theory is also useful for evaluating uncertainty (estimation/prediction error) of ground motion prediction

Thank You for Your Attention !!

Questions ??

- 請問地震風險評估是運用哪些統計方面的能力呢?
- 請問地震風險評估的統計模型與分析是指在板塊接觸地震帶所建立的建築及任何土木工程建物等等，利用統計方法來評估其因為地震而可能造成的損失，進而做安全性評估及建材使用嗎?

A:

- 地震風險評估包含了兩個主要的分析項目，一個是地震危害度分析，該分析主要為了得到目標項目所在地地震動大小的機率預測，另一個是易損性分析，該分析主要為了得到評估目標項目（ex.結構物）在某個地震動大小時不同損壞狀況的機率，經過這兩個項目的分析，我們就可以得到地震對評估項目的風險，而這兩個項目的分析過程都需要使用到適當的機率模型與統計迴歸分析技術。

- 教授的paper有沒有辦法成為一項準則，使受到地震損傷較大的地區follow，把傷亡降到最低呢?
- 地震與我們的生命安全息息相關，統計是如何運用在地震相關的模型上?

A:

- 統計在地震相關模型上的應用相當的廣泛，包括有地震的發生時序模型、發震規模大小的機率模型、發生位置的機率模型、地震動強度的機率模型等。本演講內容是地震動強度的機率模型，屬於地震危害度評估，而建物結構易損性評估 (地震會造成結構的年損壞機率) 需要依據地震危害度分析的結果進行，是整個結構耐震評估程序中相當前端與重要的工作

- … 地層底下的活動之繁複，地層下的活動、斷層的位移真的那麼好被預測以及模擬嗎?台灣蓋一些設施(建築物、橋)時真的有把地震風險實質納入考量嗎?政府是否會因為財團的金錢誘惑、民眾貪圖方便(ex明明不適合但是多了一座橋、一個隧道可以更方便)而放寬安全指標?

A:

- (a) 國內有許多專家學者透過地質調查、地球物理研究的方法，致力於斷層的探勘與研究，但仍有部分的斷層無法在其錯動之前探測到它的存在。

- (b) 台灣的建築與橋樑結構於設計階段皆須遵照現行之“耐震設計規範”來進行設計，而耐震設計規範的訂定則是依據地震危害度分析與結構易損性分析的結果，評估地震所可能造成結構的風險，並以『小震不壞、大震不倒』的性能要求，訂定合理的結構強度作為設計的準則。

- 目前地震預測的技術逐漸在進步當中，像是很多人在地震前就會收到簡訊提醒大家要做好自身的防護措施，那目前地震風險評估的模型面臨最大的挑戰是什麼?有什麼是現階段模型沒有辦法考慮的因素?
- 台灣目前的地震預測系統目前的實行狀況為何？能不能利用相關的概念對於社會大眾設計類似的預警系統等？
- 台灣的地震風險評估實際上是如何操作呢?目前是否有模型可以預測地震的發?

A:

- 所謂的地震預測指的是在地震發生前預估發生的時間、地點與大小，而地震預警指的是地震發生後在震波到達之前發布警告訊息。現階段沒有任何技術能夠做準確的地震預測，但地震預警則是可以實現的，尤其在距離震央較遠的地區能夠爭取到較長的預警時間，該技術在日本已被廣泛的應用於民眾的日常生活之中。在台灣，目前中央氣象局與國家地震工程研究中心分別有發展區域型與現地型的地震預警系統，惟台灣面積較小，震央距離都市的距離較短，因此可爭取的預警時間較短，如何爭取更長的預警時間則是目前最大的挑戰。

- 地震風險評估則是在分析地震所可能造成評估對象的損失狀況，包含了兩個主要的分析項目，一個是地震危害度分析，另一個是易損性分析，經過這兩個項目的分析，我們就可以得到地震對評估項目的風險。地震風險評估最大的挑戰在於建立精確的統計迴歸模型，而這需要大量地震與其他的觀測與分析資料，以地震觀測資料為例，目前台灣觀測地震的歷史僅有數十年，如何利用有限的觀測地震資料來建置相關的模型則是目前最大的挑戰。

- 若是想要評估未開發國家的地震風險，但手中有的資料甚少，如何建置完整的地理資訊系統資料庫來蒐集統計資料？
- 在地震頻繁的國家中做地震的風險評估統計與地震不頻繁的國家預測結果誤差上差異是否很大?
- 請問評估地震風險的話，是包含無感地震與有感地震，還是單就有感地震而已？

A:

- 進行地震風險評估需要許多不同類型的資料，包括當地的地體構造、地質構造、過去觀測的地震歷史資料、地震動的資料、評估對象的耐震能力資訊等等，資料蒐集的越齊全地震風險評估的結果才會越準確。
- 地震對評估目標物的風險，除了與地震發生的頻率與地震動的特徵有關，也與評估目標物的耐震能力相關，一般來說在評估目標物耐震能力接近的情況，地震頻繁的國家其地震所造成的風險較地震不頻繁的國家來的高。
- 針對一般結構物通常只需針對有感地震評估風險，但對例如高科技廠房等對於震動較為敏感的評估對象，則亦須考慮無感地震對其可能造成的衝擊。