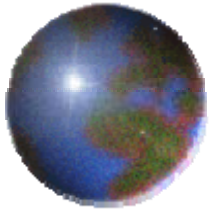
4nd Taiwan-Japan Joint Workshop on Hydrological Research for Earthquake Prediction



Outlier Detection for Anomaly Groundwater Level Time Series

Tzong-Yeang Lee, Shu-Chen Lin, Feng-Sheng Chiu,

Youe-Ping Lee, and Chi-Cheng Yang

Sep. 13-14, 2005

Acknowledgement

This work was supported in part by the Water Resources Agency (WRA).

The authors would like to thank the Disaster Protection Research Center (DPRC) for kindly permitting us to participate the “Planning of Groundwater Anomalies Associated with the Earthquake” project.

AGENDA

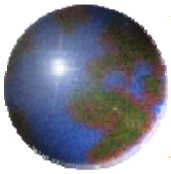
- Motive and Purpose
- Introduction
- Method and Procedure
  Outlier Analysis (OA)
  Anomaly Announcement Form (AAF)
  Factor and/or Noise Filtering
- Case Studies and Concluding Remarks
- Suggestions

Motive and Purpose

- To explore the anomaly time and pattern of seismic groundwater level (GWL) by the objective and quantitative method (outlier analysis, OA) and the specific function may be used to explain the transfer mechanism.
- The results of OA are used to assist or support the anomaly announcement form (AAF) and the suggested threshold value are also proposed for the practical application.

Introduction

- The anomaly phenomenon take place frequently before and after the earthquake, such as changes in magnetic field, ionosphere, crust strain, and groundwater. Among which, the groundwater is recorded to be highly sensitive to crust strain in many studies; it could detect slight crust strain (10^-8 volume strain).



Introduction

- The groundwater is apt to receive influences of the environmental factors or noises, like as rainfall, tide, atmospheric pressure, river water-level and artificial pumping. It increases the difficulty to analyze the variability of groundwater induced by the earthquake.

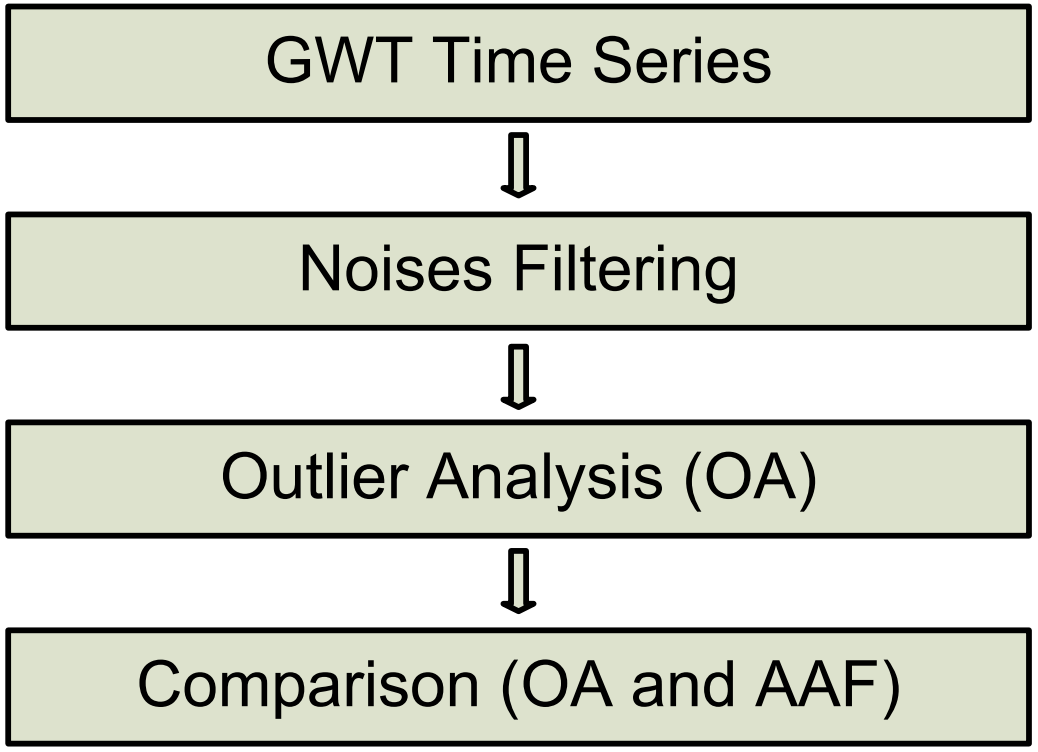
Introduction

- To analyze these effects objectively, the noises to affect the groundwater must be filtered out in advance. This purpose is convenient for analysis and the interpretation of phenomena.
- The BAYTAP-G model developed from Japan is selected for noises filtering.

Introduction

- For the purpose of anomaly detection, the outlier analysis (OA) in statistics is selected for the study.
- The OA method is used to analysis the anomaly of GWL data series after noises filtering.

Flowchart of Data Analysis



Outlier Analysis (OA)

- Time series observations are sometimes influenced by interruptive, unexpected, uncontrolled events, or even unnoticed errors of typing and recording. The consquences of these interruptive events create spurious observations that are inconsistent with the rest of time series. Such observations are usually referred to as ***outliers***.

Outlier Analysis (OA)

- When the timing and causes of interruptions are known, their effects can be accounted for by using the intervention analysis (IA) discussed in previus studies.
- The opportunity to use the IA:
  - The starting-point of intervention event is clear。
  - Specify the possible pattern of intervention impact.

Outlier Analysis (OA)

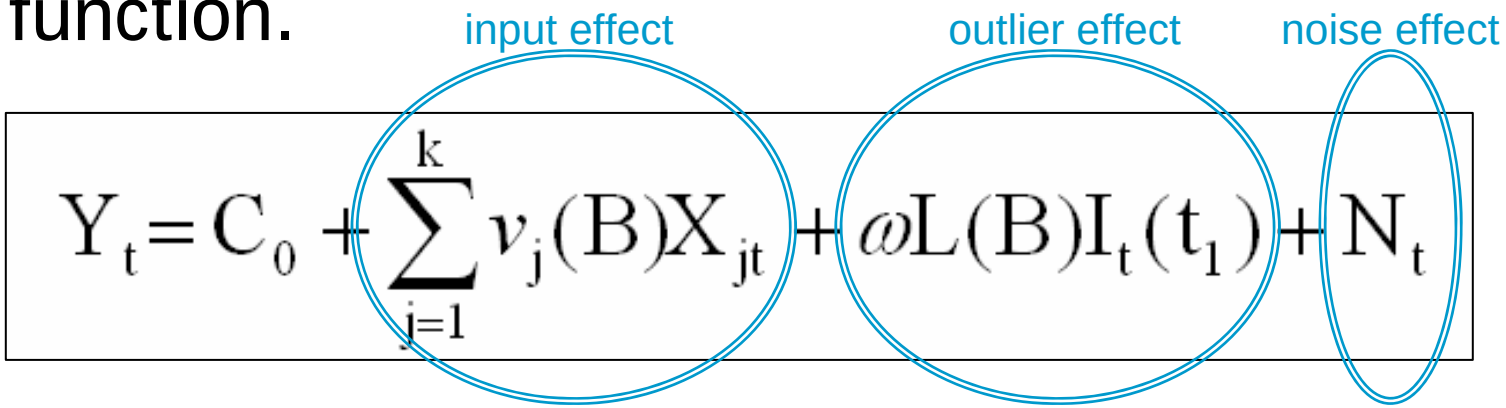
- In practice, howerver, the timing of interruptive events are usually unknown. Because outliers are known to make the resultant inference unreliable or even invalid, it is important to have procedures, like as the outlier analysis (OA) in this study, that will detect and remove such outliers effects.
- The detection of time series outliers was first studied by Fox (1972).
- The main reference in this study is Chen et al., (1990).

Outlier Analysis (OA)

- It is important to detect outliers for a number of reasons:
  - Better understanding of the series under study.
  - Better modeling and estimation.
  - Improved intervention analysis.
  - Better forecasting performance.

Outlier Analysis (OA)

- The full equation of modeling the effects of outliers includes:
  1. modeling the noise effects by ARIMA.
  2. modeling the input effects by dynamic regression.
  3. modeling the outlier effects by specific function.



Outlier Analysis (OA)

$$Y_t = C_0 + \sum_{j=1}^{k} \nu_j(B) X_{jt} + \omega L(B) I_t(t_1) + N_t$$

Y_t	output variable at time t
C_0	constant
X_{jt}	input variable j at time t
$\nu_j(B)$	polynomial of B express the relationship of Y_t and X_{jt}
ω	appear the preliminary effect of outlier or noise to output Y_t
$L(B)$	polynomial of B appear the dynamic response of outlier or noise to output Y_t
$I_t(t_1)$	indicator variable at time t_1 $I_t(t_1)$ assumes the value 1 when $t = t_1$ and is 0 otherwise
N_t	noise variable at time t assumed to follow the ARIMA model

Outlier Analysis (OA)

- Four types and L(B) functions of outliers:

Additive Outlier (AO): $L(B) = 1$

Innovational Outlier (IO): $L(B) = \frac{\theta(B)}{(1-B)^d \phi(B)}$

Level Shift (LS): $L(B) = \frac{1}{(1-B)}$

Temporary Change (TC): $L(B) = \frac{1}{(1-\delta B)}$ $0 < \delta < 1$

Outlier Analysis (OA)

- Additive Outlier (AO)
  An AO is an event that affects a series for one time period only. One illustration of an AO is a recording error.
- Innovational Outlier (IO)
  An IO is an event whose effect is propagated accroding to the ARIMA model of the process. In this manner, an IO affects all values observed after its occurrence. An IO often represents the onset of an external cause.

Outlier Analysis (OA)

- Level Shift (LS)
  A LS is an event that affects a series at a given time, and whose effect becomes permanent. A LS could reflect the change of a process mechanism, the change in recording device, or a change in the definition of the variable itself.
- Temporary Change (TC)
  A TC is an event having such an initial impact and whose effect decays exponentially accroding to some dampening factor, say δ.

Outlier Analysis (OA)

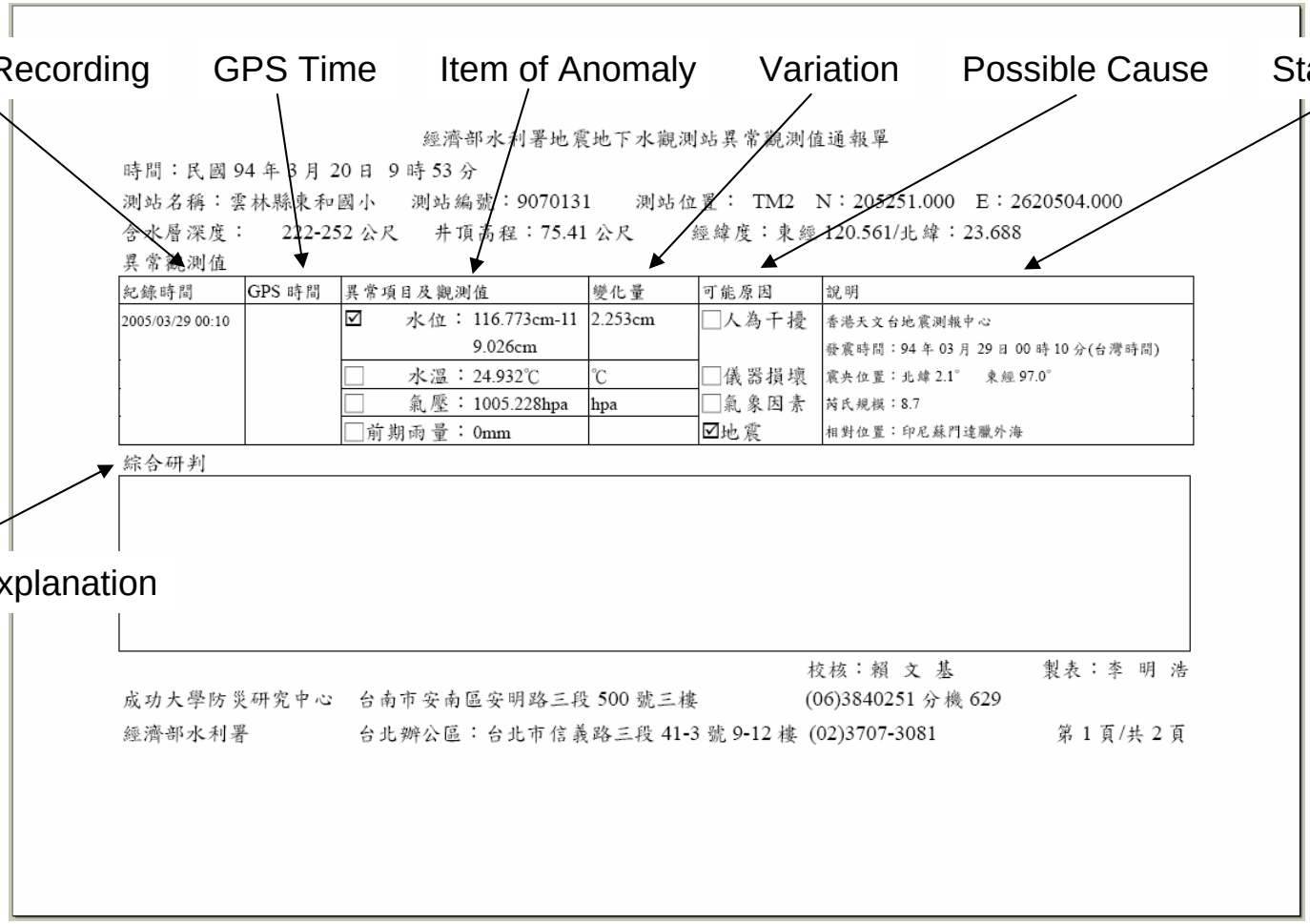
- Chen and Liu (1990) propose an iterative procedure for the joint estimation of model prarmeters and outlier effects.
- This procedure provides the basis of the SCA (scientific computing associates) software for the estimation of a time series model in the presence of possible outliers.

Anomaly Announcement Form (AAF)

- The control and management procedure of data from the groundwater observation wells in this project is to go on according to the following several steps:
  (1) Measurement of environmental information
  (2) Recording/Storage of environmental information
  (3) Checking and processing of environmental information
  (4) Noise filtering and data analysis ← ***By BAYTAP-G Model***
  (5) Identification/Determination of anomaly
  (6) Data explanation and anomaly description
  (7) Making and proposing of form

An Example of AAF (1/2)

Time of Recording | GPS Time | Item of Anomaly | Variation | Possible Cause | Statement



經濟部水利署地震地下水觀測站異常觀測值通報單

時間：民國 94 年 3 月 20 日 9 時 53 分

測站名稱：雲林縣東和國小　測站編號：9070131　測站位置： TM2 N：205251.000 E：2620504.000

含水層深度： 222-252 公尺　井頂高程：75.41 公尺　經緯度：東經 120.561/北緯：23.688

異常觀測值

紀錄時間	GPS 時間	異常項目及觀測值	變化量	可能原因	說明
2005/03/29 00:10		☑ 水位：116.773cm-119.026cm	2.253cm	☐人為干擾	香港天文台地震測報中心
		☐ 水溫：24.932℃	℃	☐儀器損壞	發震時間：94 年 03 月 29 日 00 時 10 分(台灣時間)
		☐ 氣壓：1005.228hpa	hpa	☐氣象因素	震央位置：北緯 2.1° 東經 97.0°
		☐前期雨量：0mm		☑地震	芮氏規模：8.7
					相對位置：印尼蘇門達臘外海

綜合研判

Integrated Explanation

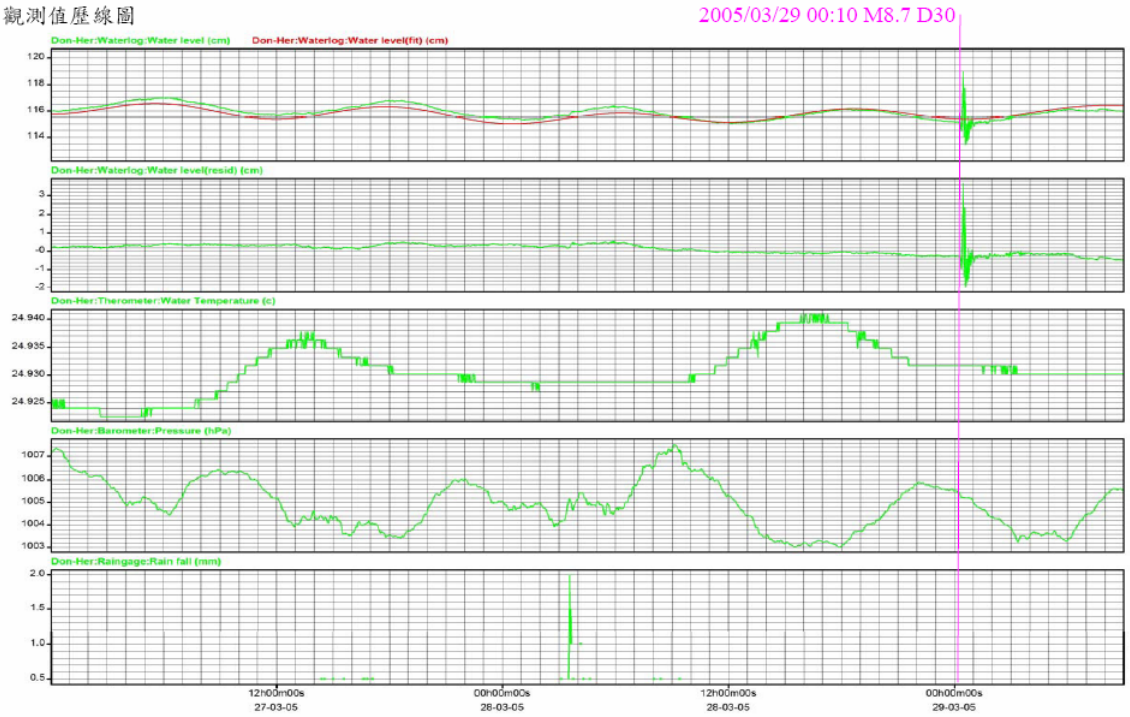
校核：賴 文 基　製表：李 明 浩

成功大學防災研究中心　台南市安南區安明路三段 500 號三樓　(06)3840251 分機 629

經濟部水利署　台北辦公區：台北市信義路三段 41-3 號 9-12 樓　(02)3707-3081　第 1 頁/共 2 頁

An Example of AAF (2/2)

Time Series Plot (TSP) of Observation



Original Groundwater Level

Residual of Groundwater Level

Temperature of Groundwater

Pressure

Rainfall

Factor/Noise Filtering

- The BAYTAP-G model is to filter the influences of affecting the GWL, including the atmospheric pressure, tide and irregular signal, etc. through the step-by-step removing. The main purpose is to appear the micro-behavior of GWL series.
- The BAYTAP-G model is adopted by this project, detail please consult earlier stage reports.

Case Studies

- OA for Real Cases - Part (I)
- Suggested OA Threshold Value of Observation Wells – Part (II)
- Comparison of OA and AAF - Part (III)

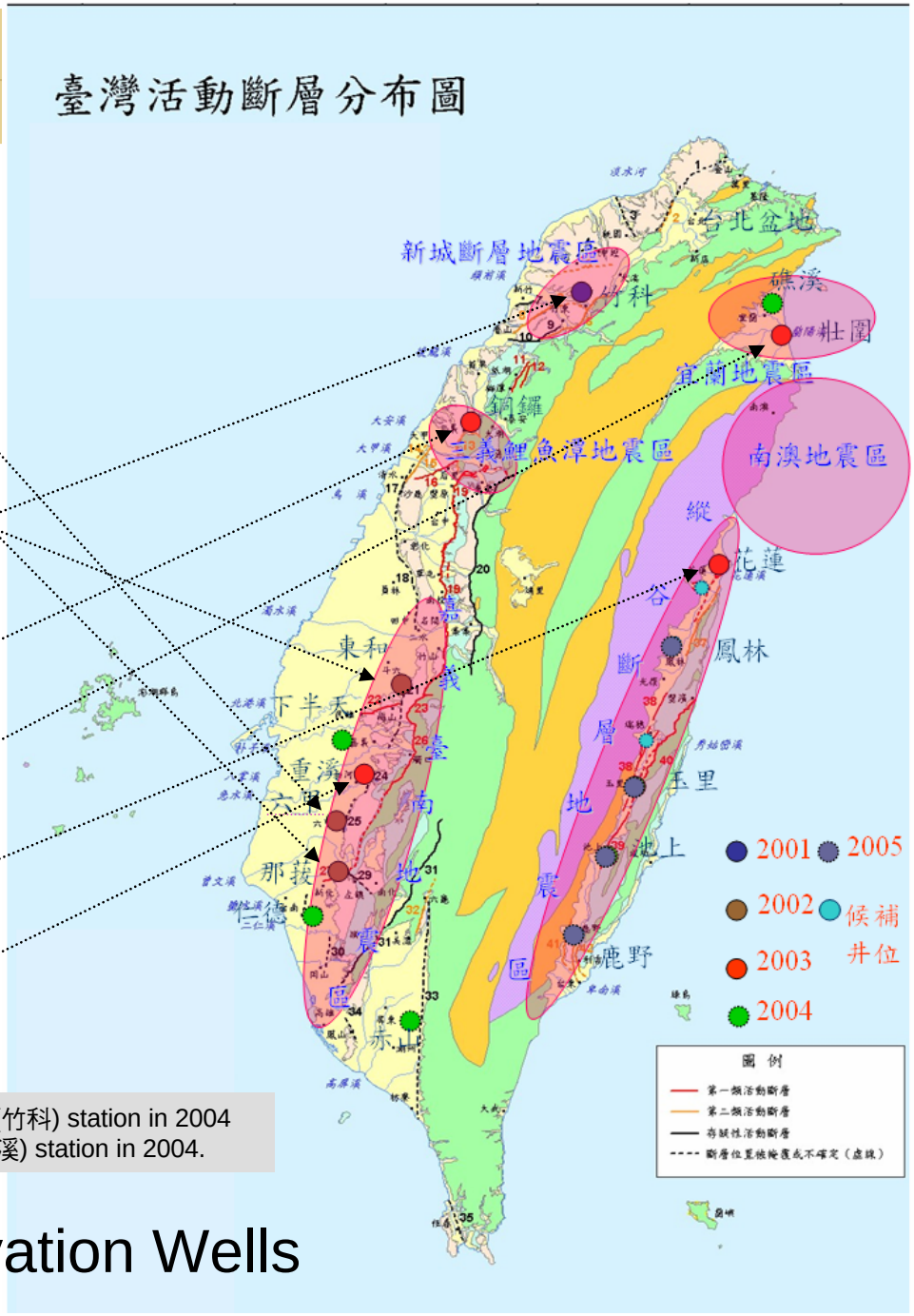
Acquistion of Data

- The data sources come from the observation stations of Water Resource Agency, Ministry of Economic Affairs (the title of project: Planning of Groundwater Anomalies Associated with the Earthquake).
- There are 8 observation wells for the study. The sampling frequency of GWL and water temperature is 2 minutes. Furthermore, the sampling frequency of atmospheric pressure and rainfall is 10 minutes.

Research Scope

- Region
  - Taiwan
  - 8 Observation Wells
- Data
  - September, 2003 ~ May, 2004
  - Data (Groundwater Level) Recording by Hourly Time Interval
  - Data Filtering by BAYTAP-G Model.

1. Liu-Jar (六甲)
2. Na-Ba (那菝)
3. Don-Her (東和)
4. Sin-Pu (新埔)
5. Tong-Lou (銅鑼)
6. Tung-Wei (壯圍)
7. Hua-Lien (花蓮)
8. Her-Don (河東)



[Notes] The Sin-Pu (新埔) station is moved to the Chu-Ker (竹科) station in 2004 and the Her-Don (河東) station is moved to the Tung-Si (重溪) station in 2004.

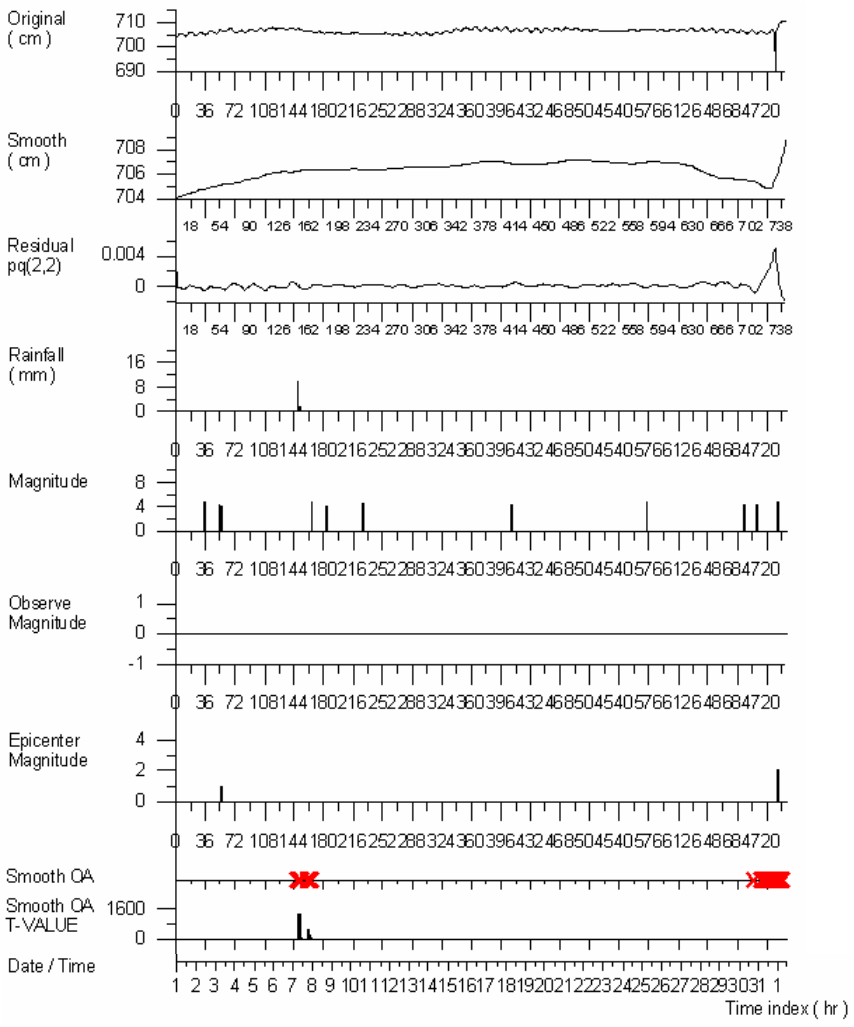
Location Map of 8 Observation Wells

Part (I) - OA for Real Cases

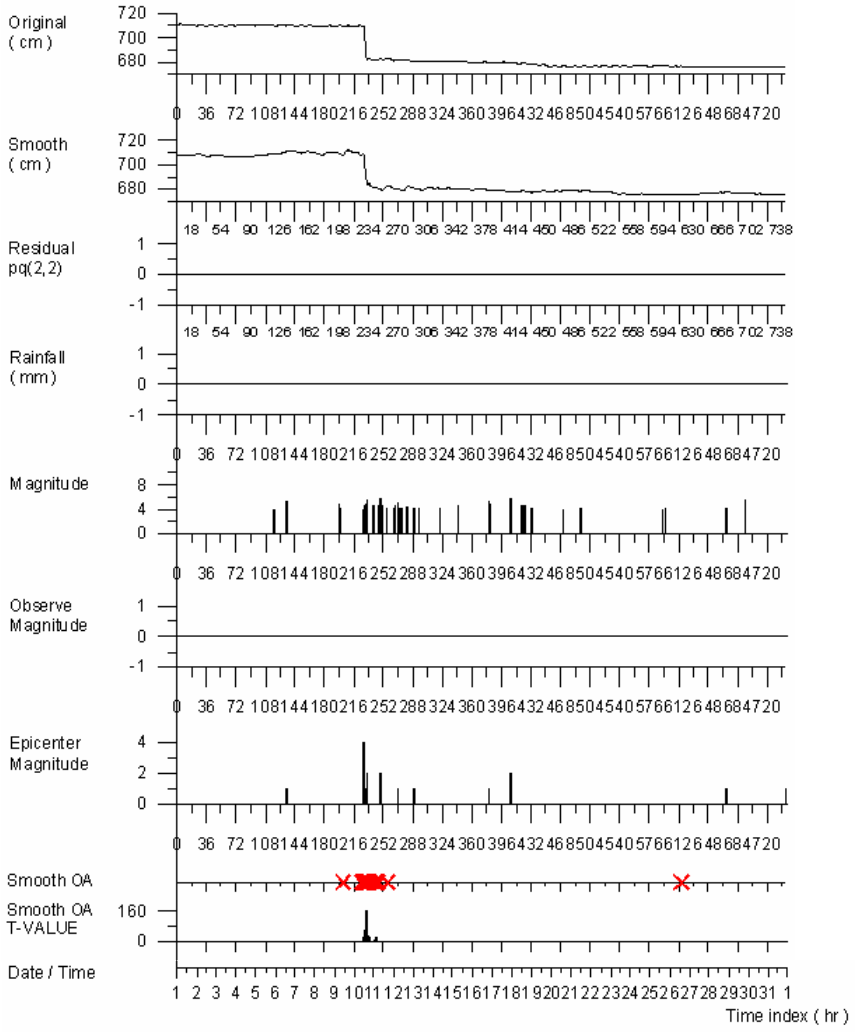
Background Statement of 12 Cases Used for Anomaly Detection by OA

Case	Station Name	Time Period for Analysis	Anomaly Event	Notes
C1	Liu-Jar (六甲)	03/01~03/31/'03	Artifical Determination	744 in total
C2	Liu-Jar (六甲)	12/01~12/31/'03	AFF(Earthquake)	744 in total
C3	Na-Ba (那菝)	06/01~06/30/'03	AFF(Earthquake)	720 in total
C4	Na-Ba (那菝)	09/01~09/30/'03	AFF(Earthquake; Rainfall)	720 in total
C5	Don-Her (東和)	09/01~09/30/'03	AFF(Earthquake; Rainfall)	720 in total
C6	Sin-Pu (新埔)	06/01~06/30/'03	Artifical Determination	720 in total
C7	Sin-Pu (新埔)	04/01~04/30/'04	AFF(Artificial Disturbance)	720 in total
C8	Hua-Lien (花蓮)	02/01~02/30/'04	AFF(Earthquake)	720 in total
C9	Hua-Lien (花蓮)	04/01~04/30/'04	AFF(Earthquake)	720 in total
C10	Tong-Lou (銅鑼)	03/01~03/30/'04	Artifical Determination	720 in total
C11	Tung-Wei (壯圍)	05/01~05/31/'04	Artifical Determination	744 in total
C12	Her-Don (河東)	04/01~04/30/'04	AFF(Artificial Disturbance)	720 in total

Part (I) - OA for Real Cases



Time Series Plot of OA in CASE C1 Original GWL Considered from the BAYTAP-G Filtering



Time Series Plot of OA in CASE C2 Original GWL Considered from the BAYTAP-G Filtering

Concluding Remarks

- The OA method has the properties of rigorous theory and complete procedure. It can be used to detect the known or unknown interruptive event. For the need and purpose of this project, affecting factors or noises should be filtered out if possible. It is expected that the result of OA can be “clear” to reflect the influence of earthquake event on the groundwater level.

Concluding Remarks

- The anomaly pattern of GWL caused by earthquake and rainfall is mainly the innovational outlier (IO). When the earthquake and rainfall coexist, the OA is not easy to distinguish between them. If the rainfall can be filtered out from the BAYTAP-G model, the OA may be more suitable and easy to explain.

Part (II) - Suggested OA Threshold Value

Suggested OA Threshold Vaule of 12 cases

Case	Station Name	Suggested Value for Individual Case	Suggested Value for Individual Station (Well)
C1	Liu-Jar (六甲)	20	20
C2	Liu-Jar (六甲)	20	
C3	Na-Ba (那菝)	10	8
C4	Na-Ba (那菝)	8	
C5	Don-Her (東和)	8	8
C6	Sin-Pu (新埔)	4	4
C7	Sin-Pu (新埔)	4	
C8	Hua-Lien (花蓮)	4	4
C9	Hua-Lien (花蓮)	4	
C10	Tong-Lou (銅鑼)	10	10
C11	Tung-Wei (壯圍)	10	10
C12	Her-Don (河東)	4	4

Concluding Remarks

- By the OA results of 8 observation stations, it is found that the number of anomaly seems partial more with the statistical test of 95% confidence limit. If we increase the testing value appropriately, it can contribute to the explanation of the anomaly. In this study, the preliminary suggestion value of 8 observation stations is provided.

Part (III) - Comparison of OA and AAF

Summary Statement of AAF of 8 Observation Wells

Date of AAF (Year/Month)	Station Name of Observation Well							
	Liu-Jar (六甲)	Na-Ba (那菝)	Don-Her (東和)	Sin-Pu (新埔)	Tong-Lou (銅鑼)	Tung-Wei (壯圍)	Hua-Lien (花蓮)	Her-Don (河東)
2003/09	—	YES (★/▼)	—	—	—	—	—	—
2003/10	—	NO	—	—	—	—	—	—
2003/11	NO	NO	NO	YES (＃)	—	—	—	—
2003/12	YES (★)	NO	YES (★)	NO	YES (★)	—	YES (★)	YES (★)
2004/01	YES (★)	YES (＃)	NO	NO	NO	YES (★)	YES (★)	YES (＃)
2004/02	NO	YES (＃)	NO	YES (＃)	YES (＃)	NO	YES (★)	YES (＃)
2004/03	NO	YES (＃)	NO	YES (▼)	YES (＃)	YES (＃)	NO	YES (＃)
2004/04	NO	YES (＃)	YES (▼)	YES (＃)	YES (▼)	NO	YES (★)	YES (＃)
2004/05	NO	YES (▼)	YES (★/▼)	YES (＃)	YES (＃)	YES (▼)	NO	YES (＃)
Symbol Statement	Not available (—); Anomaly in AAF (YES); No anomaly in AAF (NO) Artificial disturbance (＃) Instrument damage (●) The meteorological factor, such as rainfall, etc. (▼) Earthquake (★) Unknown reason (？)							

[Notes] The Sin-Pu (新埔) station is moved to the Chu-Ker (竹科) station and the Her-Don (河東) station is moved to the Tung-Si (重溪) station.

Part (III) - Comparison of OA and AAF

Comparison of OA and AAF in Liu-Jar (六甲) Station

Station Name Liu-Jar(六甲)	AAF	OA	
		Time-Point	Pattern
11/2003	NO	--	--
12/2003	228	T=226 T=228 T=229 T=230	IO IO IO IO
	229	T=231	IO
	230	T=231 T=232	IO IO
01/2004	336 to 600	--	--
02/2004	NO	--	--
03/2004	NO	--	--
04/2004	NO	--	--
05/2004	NO	--	--

[Notes] The symbol "--" means that not any anomaly is detected by OA.

Comparison of OA and AAF in Na-Ba (那菠) Station

Station Name Na-Ba(那菠)	AAF	OA	
		Time-Point	Pattern
09/2003	604	--	--
10/2003	NO	--	--
11/2003	NO	--	--
12/2003	NO	--	--
01/2004	247	T=247 T=248	IO LS
	257	T=257 T=258	IO TC
	277	T=277 T=278	IO LS
02/2004	59	T=60	LS
	65	T=65 T=67	IO IO
03/2004	398	T=400	LS
	415	T=417	IO
	491	T=493	IO
04/2004	715	T=715 T=716 T=717	IO IO IO
05/2004	199	T=200	LS
	471	T=471 T=472	IO LS

[Notes] The symbol "--" means that not any anomaly is detected by OA.

Concluding Remarks

- To compare the results of OA to the AAF, the success ratio is near 82%. The AAF with seven-steps procedure is moderately subjective, but the OA with the standard operation procedure is more objective.
- The OA can be used for the automatic detection of anomaly time-point, the automatic detection of possible interruptive event, and the automatic arrangement of anomaly pattern. These properties can not be adequately provided by the current procedure of the AAF.

Suggestions

- Further studies are focused on the topics as follows:
  It seems to be sensitive that adopts the 95% confidence limit to parameter testing of the anomaly. We suggest that the suitable threshold value can be analyzed and adjusted based on the historical record of every observation well.
  The effect of rainfall is still not filtered out by the BAYTAP-G model. It causes some difficulties of interpretation, so the filtering of rainfall data should be necessary.

THANKS FOR
YOUR
ATTENTION AND COOPERATION