

# OLS 및 변수선택법에 의한 다중선형회귀모형 매개변수 산정

## Parameter Estimation for Multiple Linear Regression Model by OLS and Stepwise

김경탁\*·김주훈\*·박정솔\*\*

Kyung Tak Kim\*·Joo Hun KIM\*·Jung Sool Park\*\*

### 요 지

본 연구는 OLS 및 변수선택법에 의해 통계학적 모형의 매개변수를 산정하여 모형의 적용성을 입증하고 하천 주요지점에 대한 홍수위 예측을 통해 홍수예보 및 예측 업무에 기여코자하는데 연구목적이 있다. 다중선형회귀모형을 구성하기 위한 독립변수는 예보지점의 수위/유출량 자료와 상류지점의 수위/유출량 자료, 그리고 유역의 선행 평균강우량 등의 자료를 독립변수로 하여 통계학적 홍수예측을 위한 다중선형 회귀모형을 각각 구성하여 적합성 여부를 판단하였다. 매개변수 산정은 OLS(Ordinary least square root method)와 변수선택(Stepwise)방법에 의해 산정하였으며, 중랑천 유역의 2002년부터 2005년까지의 수문사상 16개를 선정하여 모형에 적용한 결과 두 매개변수 산정방법 모두 30분에서 90분 예측은 상대적으로 정확한 결과를 나타내었으며, OLS 및 변수선택법에 의한 매개변수 산정결과 변수선택법에 의한 방법이 OLS 방법보다는 상관성이나 효율지수면에서 조금 더 정확한 값을 나타내고 있으나 독립변수의 일관성을 감안한다면 변수선택법보다는 OLS방법에 의한 매개변수 산정이 타당할 것으로 사료된다. 기존의 홍수예보 업무에 활용되고 있는 수문학적 홍수예측 모형인 저류함수법의 여러 매개변수 조정에 의한 홍수위 예측 방법보다는 비교적 간단한 통계적 방법에 의한 홍수위 예측 방법으로 홍수예보의 선행시간 확보가 필수적인 중랑천과 같이 유역면적이 작은 중소하천에서의 홍수예보 업무에 효과적으로 이용 가능할 것으로 사료된다.

### 1. 서 론

하천의 중요지점에 대한 홍수예보 및 예측업무를 수행하기 위해서는 유역으로부터의 유출현상을 예측해야 하는데 이는 기본적으로 관측 또는 예측된 강우 입력자료를 이용하여 유량을 예측하는 강우-유출 분석을 수행하여야 한다. 대부분의 강우-유출 모형은 사용되는 자료와 분석 목적에 따라 다르겠지만 비선형성이 강하고 유역의 지형학적 인자와 기후학적 인자의 영향을 포함하기 때문에 매우 불규칙하여 정확한 예측이 어렵다.

현재 국내 주요 하천의 홍수예보시스템의 운영과 다목적댐의 홍수조절관리에는 수문학적 모형의 하나인 저류함수모형(Storage Function Model, SFM)이 사용되고 있다. 저류함수모형을 실제 홍수유출 현상에 적용하는데 있어서 가장 어려운 점은 매개변수를 결정하는 것이며, 특히 유역면적이 작아 홍수 도달시간이 짧은 유역에서의 수문학적 모형에 의한 홍수예보 업무는 더더욱 어려운 것이 현실이다.

본 연구에서는 유역면적이 작아 홍수예보 발령시간을 충분히 확보하지 못하는 한강수계의 중랑천 유역을 대상으로 강우-유출 모형이 요구하는 광범위한 경계조건 및 초기조건 자료가 필요없이 유역의 중요지점에 대한 홍수량을 예측할 수 있는 통계학적 모형을 개발하여 홍수예보 및 예측 업무에 활용하는 것을 목적으로 하였다.

\* 정회원·한국건설기술연구원 수자원연구부 선임연구원·E-mail:ktkim1@kict.re.kr

\* 정회원·한국건설기술연구원 수자원연구부 선임연구원·E-mail:jh-kim@kict.re.kr

\*\* 정회원·한국건설기술연구원 수자원연구부 연구원·E-mail:parkjs@kict.re.kr

## 2. 홍수예측 모형

### 2.1 대표매개변수 선정

본 연구의 대상유역인 중랑천 유역의 경우 대부분의 강우/수위관측소가 2002년부터 관측을 개시하여 홍수예보 업무에 활용하고 있다. 본 연구에서는 이들 유역의 우량관측소 관측 자료 중 80mm 이상의 호우사상을 수집하여 통계학적 모형의 매개변수 선정 및 모형의 적합성 여부를 검증하였다.

통계학적 홍수예측 모형 개발을 위한 호우사상 선정은 호우 원인별로 기압골과 전선형, 태풍 등을 종합하여 첨두 홍수량이 크고 지속기간이 짧은 호우 사상을 선정하였으며, 2002/07/05 ~ 07/07의 전선형 호우사상을 선정하여 통계학적 매개변수 선정에 이용하였다.

통계학적 홍수예측 모형의 독립변수는 다음과 같이 예보지점의 수위/유출량 자료와 상류지점의 수위/유출량 자료, 그리고 유역의 선행 평균강우량 등의 자료를 독립변수로 하여 통계학적 홍수예측을 위한 다중선형 회귀모형을 각각 구성하여 독립변수의 적합성을 판단하였다.

- 홍수예보지점의 세시간 이전 유역 평균강우량 자료와 유출량 자료
- 홍수예보지점의 세시간 이전 유역 평균강우량 자료와 수위 자료
- 홍수예보지점의 세시간 이전 유출량 자료
- 세시간 이전 본류 및 지류의 상류지점 수위, 홍수예보지점 수위, 평균강우량 자료

상기와 같이 독립변수를 구성하여 분석한 결과 세 시간 이전의 유역 평균강우량자료와 예보지점의 유출량 자료를 이용하여 다중선형회귀 홍수예측 모형을 구성하였을 때 가장 좋은 결과를 나타내었다.

표 1. 독립변수 선정을 위한 통계분석

독립변수	상관계수 (CC)	상대첨두오차 (Peak relative error)	효율지수 (EI)
1) 유출량 자료 이용			
유출량자료(중랑교, 신곡, 우이)를 이용한 분석(OLS)	0.99	0.19	0.87
유출량자료(중랑교, 신곡)를 이용한 분석(OLS)	0.99	0.33	0.74
유출량자료(중랑교, 우이)를 이용한 분석(OLS)	0.99	0.12	0.93
2) 수위 자료 이용			
수위자료(중랑교, 신곡, 우이)를 이용한 분석(OLS)	0.97	0.88	-6.95
수위자료(중랑교, 신곡)를 이용한 분석(OLS)	0.97	1.16	-10.38
수위자료(중랑교, 우이)를 이용한 분석(OLS)	0.99	0.91	-10.76
3) 강우-수위 자료 이용			
강우-수위자료(중랑교, 신곡, 우이)를 이용한 분석(OLS)	0.99	0.04	0.95
강우-수위자료(중랑교)를 이용한 분석(OLS)	1.00	0.07	0.98
강우-수위자료(중랑교)를 이용한 분석(Stepwise)	1.00	0.05	0.99
4) 강우-유출량 자료 이용			
강우-유출량자료(중랑교)를 이용한 분석(OLS)	0.99	0.02	0.97
강우-유출량자료(중랑교)를 이용한 분석(Stepwise)	0.99	0.00	0.98

### 2.2 홍수예측 모형

유역 홍수의 실시간 예측을 위해서 다중선형회귀 홍수예측 모형은 실시간으로 수집되는 자료와 종류를 고려하여 모형을 그림(1)과 같이 구성하였고, 그림(1)에서 수위예측 모형은  $t - m$  시간까지의 유역 평균 선행 강우량과  $t - m$  시간까지의 선행유출량을 입력자료로 하며, 모형의 출력으로는  $t + n$  시간까지의 수위예측이 가능하다. 따라서 대상지점에서의 선행시간을 3시간 ( $n = 6$ ), 강우량 및 유출량의 선행 지체시간을 2.5시간 ( $m = 5$ )으로 설정하였다. 또한  $t - 5$  시간까지의 선행 강우량 및 선행 유출량의 입력에 의한 모형의 전달함수는 선형으로 가정하였다.

출력값인 홍수량은 예보지점에서의 수위-유량 관계곡선식에 의해 수위값으로 환산되어 홍수예보 업무에 활용될 수 있으며, 입력자료에 따른 모형을 수식화하면 식(1)과 같이 표현할 수 있다.

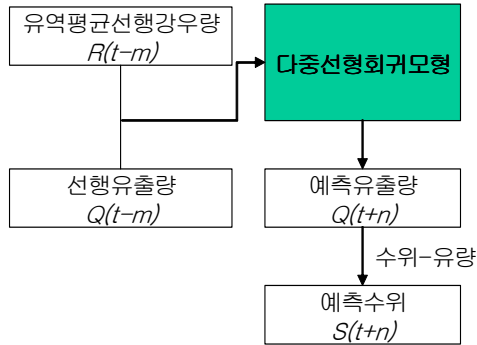


그림 1. 홍수량 예측모형의 구조

모형의 매개변수를 표(3)과 같이 산정하였다.

$$Q(t+n) = a_1Q(t) + a_2Q(t-1) + \dots + a_{m+1}Q(t-m) + b_1R(t) + b_2R(t-1) + \dots + b_{n+1}R(t-m) + c$$

(1)

여기서,  $Q$ 는 유출량,  $R$ 은 대상지점 상류유역의 평균강우량,  $a_1, a_2, \dots, a_{m+1}$ 은 대상지점의 유출량 관련 매개변수,  $b_1, b_2, \dots, b_{n+1}$ 는 대상지점의 유역평균 강우량 관련 매개변수,  $c$ 는 절편매개변수이며,  $n$ 은 선행예측시간을 나타낸다.

유역의 3시간 이전까지의 평균 강우량과 유출량 자료를 이용하여 통계모형을 구성한 매개변수에 대한 적합성 분석 결과 표(2)와 같이 매우 우수한 결과를 나타내었으며, 이 결과에 대한 통계학적 홍수예측

표 2. 통계모형 매개변수 선정 분석 결과

	30분			60분			90분		
	$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE
OLS 방법	1.00	1.00	0.86	0.99	0.98	1.74	0.97	0.97	2.53
	120분			150분			180분		
	$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE
	0.95	0.95	3.23	0.93	0.92	3.98	0.89	0.89	4.76
Stepwise 방법	30분			60분			90분		
	$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE
	1.00	1.00	0.86	0.98	0.98	1.82	0.97	0.97	2.55
	120분			150분			180분		
$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE	$r^2$	수정 $r^2$	SE	
0.95	0.95	3.22	0.92	0.92	3.97	0.89	0.89	4.80	

표 3. 통계모형 매개변수

매개변수	30분		60분		90분		120분		150분		180분	
	OLS	Stepwise	OLS	Stepwise	OLS	Stepwise	OLS	Stepwise	OLS	Stepwise	OLS	Stepwise
$c$	0.019	0.018	0.115	0.166	0.269	0.236	0.466	0.475	0.706	0.786	0.987	1.098
$a_1$	1.667	1.681	1.820	1.538	1.683	1.507	1.717	1.439	1.705	1.268	1.596	0.680
$a_2$	-0.992	-1.043	-1.438	-0.633	-1.263	-0.643	-1.393	-0.630	-1.550	-0.516	-1.635	
$a_3$	0.360	0.469	0.837		0.759		0.787		0.836		0.685	
$a_4$	0.122		-0.076		-0.034		-0.054		-0.217		-0.015	
$a_5$	-0.178	-0.130	-0.194		-0.372		-0.617		-0.563		-0.722	
$a_6$	0.000		-0.017		0.100		0.368		0.525		0.744	
$b_1$	1.333	1.336	2.557	2.036	3.885	3.568	5.008	4.878	6.065	6.150	6.912	7.680
$b_2$	-1.015	-1.047	-1.105		-1.077		-1.167		-1.319		-1.362	
$b_3$	0.823	0.866	1.257		1.334		1.464		1.649		1.875	
$b_4$	-0.393	-0.388	-0.656		-0.618		-0.658		-0.571		-0.179	
$b_5$	0.097	-0.240	0.195		0.032	-0.660	0.124	-1.093	0.435	-1.577	0.433	-1.954
$b_6$	-0.316		-0.771		-0.948		-0.953		-1.116		-1.084	

### 3. 홍수량 예측

홍수량 예측을 위한 수문자료는 2002년부터 2005년 9월까지의 강우량 중에서 80mm이상의 호우를 대상으로 적용하였다. 모형의 적용결과 그림(2) 및 그림(3)에서 보는바와 같이 30분 및 60분예측의 경우 육안분석만으로도 매우 정확한 예측을 하고 있는 것으로 판단되었다.

통계학적 홍수예측 모형의 검증은 평균제곱근오차, 상대최대오차, 효율지수, 상관계수, 평균절대편차, 그리고 절대평균편차와 평균값 사이의 비인 RAD, 그리고 홍수예측에서 가장 중요한 요소인 상대침두오차와 상대침두시간오차를 모형의 적합성 판정을 위한 통계치로 사용하였다.

적용 결과 30분예측의 경우 수문사상에 따라 다소 차이가 있으나 OLS 및 변수선택법 모두 상관계수가 0.92 ~ 1.00으로 매우 높은 상관관계를 보이고 있으며, 60분예측에서도 0.85 ~ 0.99(OLS) 및 0.89 ~ 0.99(변수선택법), 그리고 90분예측에서는 0.77 ~ 0.98(OLS), 0.81 ~ 0.99(변수선택법)로 다소 변수선택법에 의한 매개변수 선정이 더 상관성이 높은 것으로 평가되나 큰 차이를 보이지는 않았다. 또한 효율지수의 경우에도 30분예측의 경우 상관계수 분석과 마찬가지로 두 방법 모두 0.83 ~ 1.00로 매우 높은 결과를 나타내고 있다. 60분예측 및 90분예측에서 변수선택법에 의한 방법의 효율지수가 다소 높게 나타나나 큰 차이를 보이지는 않았다. 절대평균편차는 0.55에서 6.95정도이며, 침두상대오차는 작은 경우에는 0.63%, 큰 값의 경우에는 약 30% 정도 되지만 대체로 매우 양호한 결과를 나타내고 있으며, 상대침두시간 오차는 30분 예측값의 경우 매우 정확한 값을 나타내고 있으며, 통계모형에 의한 홍수예측 결과 두 매개변수 산정방법 모두 30분예측에서 90분예측까지는 비교적 정확한 결과를 나타내었다.

표 4. 통계학적 모형 검증결과(상관계수)

호우기간	항목	OLS 방법						Stepwise 방법					
		30분	60분	90분	120분	150분	180분	30분	60분	90분	120분	150분	180분
02/04/29 ~ 05/02		0.96	0.92	0.89	0.85	0.79	0.74	0.96	0.97	0.94	0.91	0.87	0.82
02/07/05 ~ 07/08		1.00	0.99	0.98	0.97	0.96	0.94	1.00	0.99	0.98	0.97	0.96	0.93
02/08/04 ~ 08/10		0.99	0.96	0.91	0.85	0.78	0.72	0.99	0.96	0.91	0.85	0.79	0.72
03/05/06 ~ 05/10		1.00	0.99	0.98	0.96	0.93	0.89	1.00	0.99	0.98	0.96	0.93	0.90
03/07/21 ~ 07/26		0.99	0.97	0.95	0.92	0.89	0.84	0.99	0.97	0.95	0.92	0.89	0.84
03/08/05 ~ 08/10		0.99	0.94	0.87	0.81	0.74	0.65	0.99	0.94	0.87	0.81	0.74	0.67
03/08/19 ~ 08/22		0.98	0.93	0.84	0.75	0.66	0.58	0.98	0.93	0.84	0.75	0.65	0.56
03/08/22 ~ 09/01		0.99	0.95	0.90	0.85	0.80	0.75	0.99	0.95	0.90	0.85	0.79	0.74
03/09/18 ~ 09/20		0.99	0.97	0.93	0.88	0.82	0.73	0.99	0.97	0.93	0.87	0.80	0.69
04/07/03 ~ 07/06		0.99	0.98	0.96	0.93	0.89	0.85	0.99	0.98	0.96	0.93	0.89	0.85
04/07/11 ~ 07/19		0.99	0.96	0.92	0.88	0.84	0.79	0.99	0.96	0.92	0.88	0.85	0.80
05/06/26 ~ 06/29		0.99	0.95	0.90	0.84	0.78	0.74	0.99	0.95	0.90	0.84	0.79	0.75
05/06/30 ~ 07/05		0.92	0.85	0.77	0.68	0.60	0.52	0.92	0.89	0.81	0.74	0.67	0.63
05/07/27 ~ 07/30		0.96	0.90	0.83	0.79	0.74	0.69	0.96	0.90	0.84	0.80	0.76	0.74
05/08/10 ~ 08/14		0.98	0.94	0.88	0.83	0.78	0.74	0.98	0.94	0.89	0.84	0.78	0.74
05/08/24 ~ 08/27		1.00	0.99	0.98	0.97	0.95	0.92	1.00	0.99	0.98	0.97	0.96	0.94
05/09/12 ~ 09/15		0.99	0.95	0.90	0.84	0.79	0.72	0.99	0.95	0.90	0.85	0.80	0.77

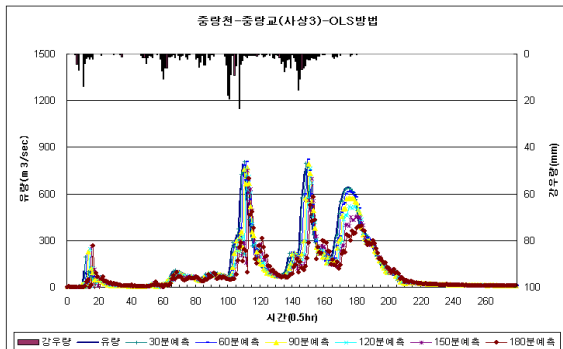


그림 2. 모형 적용 결과(OLS방법)

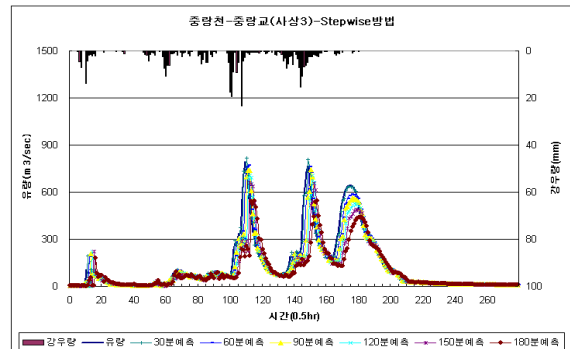


그림 3. 모형 적용 결과(Stepwise방법)

표 5. 통계학적 모형 검증결과(효율지수)

호우기간	항목	OLS 방법					Stepwise 방법						
		30분	60분	90분	120분	150분	180분	30분	60분	90분	120분	150분	180분
02/04/29 ~ 05/02		0.92	0.84	0.77	0.70	0.57	0.50	0.91	0.94	0.88	0.83	0.75	0.65
02/07/05 ~ 07/08		1.00	0.98	0.96	0.94	0.91	0.88	1.00	0.98	0.96	0.94	0.91	0.87
02/08/04 ~ 08/10		0.98	0.91	0.82	0.71	0.59	0.47	0.98	0.92	0.82	0.71	0.60	0.49
03/05/06 ~ 05/10		0.99	0.98	0.95	0.91	0.85	0.77	0.99	0.98	0.95	0.91	0.86	0.78
03/07/21 ~ 07/26		0.99	0.95	0.90	0.83	0.77	0.67	0.99	0.94	0.89	0.83	0.77	0.67
03/08/05 ~ 08/10		0.97	0.88	0.75	0.65	0.53	0.40	0.98	0.88	0.76	0.65	0.54	0.43
03/08/19 ~ 08/22		0.97	0.86	0.68	0.53	0.40	0.30	0.97	0.86	0.70	0.53	0.38	0.27
03/08/22 ~ 09/01		0.97	0.89	0.80	0.72	0.63	0.55	0.97	0.90	0.81	0.72	0.62	0.53
03/09/18 ~ 09/20		0.98	0.94	0.86	0.76	0.65	0.50	0.98	0.94	0.85	0.75	0.63	0.46
04/07/03 ~ 07/06		0.99	0.96	0.91	0.85	0.77	0.69	0.99	0.96	0.92	0.86	0.77	0.69
04/07/11 ~ 07/19		0.98	0.93	0.84	0.76	0.67	0.55	0.98	0.92	0.84	0.76	0.68	0.59
05/06/26 ~ 06/29		0.97	0.90	0.79	0.67	0.55	0.47	0.97	0.90	0.80	0.68	0.57	0.52
05/06/30 ~ 07/05		0.83	0.67	0.53	0.35	0.21	0.08	0.83	0.77	0.63	0.50	0.39	0.34
05/07/27 ~ 07/30		0.92	0.92	0.79	0.66	0.58	0.41	0.92	0.92	0.81	0.69	0.62	0.52
05/08/10 ~ 08/14		0.96	0.88	0.77	0.67	0.57	0.49	0.96	0.89	0.79	0.68	0.58	0.50
05/08/24 ~ 08/27		0.99	0.97	0.95	0.92	0.88	0.82	0.99	0.98	0.95	0.94	0.90	0.86
05/09/12 ~ 09/15		0.98	0.90	0.80	0.69	0.58	0.46	0.98	0.89	0.80	0.70	0.62	0.57

4. 결론

본 연구는 유역면적이 작아 홍수예보 발령시간을 충분히 확보하지 못하는 한강수계의 중랑천 유역을 대상으로 유역의 중요지점에 대한 홍수량을 예측할 수 있는 통계학적 모형을 개발하여 OLS 및 변수선택법에 의해 모형의 매개변수를 산정하여 두 방법간 적합성을 판단하여 모형의 적용성을 입증하고 하천 중요지점에 대한 홍수위 예측을 통해 홍수예보 업무에 활용하는 것을 목적으로 하였으며, 연구결과를 요약하면 다음과 같다.

매개변수 선정을 위한 수문자료는 2002/07/05-07/07의 전선형 호우사상을 선정하였고, 다중선형회귀모형을 구성하기 위한 독립변수는 상기의 호우사상을 대상으로 예보지점의 수위/유출량 자료와 상류지점의 수위/유출량 자료, 그리고 유역의 선행 평균강우량 등의 자료를 독립변수로 하여 통계학적 홍수예측을 위한 다중선형 회귀모형을 각각 구성하여 적합성 여부를 평가하였으며, 독립변수의 선정 결과 세시간 이전까지의 유출량과 강우량 자료를 이용하여 분석하는 것이 가장 타당한 것을 판단되었다(표(1)).

모형의 적용결과 두 매개변수 산정방법 모두 30분에서 90분 예측은 상대적으로 정확한 결과를 나타내었으며, 변수선택법에 의한 방법이 OLS 방법보다는 상관성이나 효율지수면에서 더 정확한 값을 나타내고 있으나 독립변수의 일관성을 감안한다면 변수선택법보다는 OLS방법에 의한 매개변수 산정이 타당할 것으로 사료된다.

기존의 홍수예보 및 예측 업무에 활용되고 있는 수문학적 홍수예측 모형인 저류함수법의 여러 매개변수 조정에 의한 홍수위 예측 방법보다는 비교적 간단한 통계적 방법에 의한 홍수위 예측 방법으로 홍수예보의 선행시간 확보가 필수적인 중랑천과 같이 유역면적이 작은 중소하천에서의 홍수예보 및 예측 업무에 효과적으로 이용 가능할 것으로 사료된다.

참고문헌

건설교통부/한강홍수통계소(2005). "주요지천홍수예보프로그램 개선"  
 윤강훈, 김태균(2004). "레이더 예측 강우를 이용한 다중회귀 예측모형의 적용가능성 평가", 대한토목학회 24(4b):0-5.  
 윤강훈, 신현민(1994). "한강인도교 수위와 영향인자간의 다중회귀분석에 의한 홍수위 예측 모형". 한국수문학회지 27(3):55-69  
 Davis, J.C.(1986). Statistics and Data Analysis in Geology. John Wiley & Sons, Inc., N. York.  
 Draper, N.R. and Smith, H.(1989). Applied Regression Analysis. John Wiley and Sons, Inc., New York.  
 Huynh Ngoc Phien and Nguyen Duc Anh Kha. (2003). "Flood forecasting for the upper reach of the Red River Basin, North Vietnam". <http://www.wrc.org.za>  
 Wilke, K. (1997). Ribamod: river basin modelling management and flood mitigation : Forecast Systems for Large Rivers - The River Rhin Catchment pp.105-126