

Original Article

Rogue Wave 예측을 위한 이상지수 모델

김도영[†]

홍익대학교 조선해양공학과 교수

Abnormality Index Model for Rogue Wave Prediction

Do Young Kim[†]

Professor, Department of Naval Architecture and Ocean Engineering, Hongik University, Sejong 30016, Korea

요 약

본 논문에서는 rogue wave가 포함된 Draupner 플랫폼에서 계측된 파랑계측 자료의 통계치에서 얻은 15개의 파랑통계 변수로 다중선형회귀모델을 이용하여 rogue wave의 이상지수 예측모델을 계산하였다. 예측모델과 실제 이상지수의 상관관계수는 0.984로 데이터가 회귀직선과 잘 일치하는 것을 볼 수 있었다. 자료의 재추출을 위해 LOOCV와 k-겹 교차검 증방법을 사용하였다. Rogue wave가 학습데이터 또는 평가데이터에 속한 경우에 따른 8가지 경우로 분류해서 모델의 오차를 분석하였다. 파랑 데이터 중 일부 rogue wave를 학습데이터에 포함된 경우가 평가오차가 가장 작게 나타나는 것을 볼 수 있었다. New Year Wave로 불리는 데이터가 포함된 학습데이터가 일반적으로 다른 세트의 학습데이터보다 평가오차가 작아지는 경향을 보였다.

Abstract – In this paper, multilinear regression with 15 wave parameters from Draupner wave data including rogue waves is used to model Abnormality index of rogue waves. The correlation coefficients between predicted and actual abnormality index is 0.984. For data resampling, LOOCV(Leave One Out Cross Validation) and k-fold Cross Validations are used. Errors in prediction of eight cases depending on the inclusion of rogue waves are examined. A model with training data group with certain rogue waves yields the lowest error in prediction of the abnormality index in rogue wave cases. The data set including the New Year Wave generally yields less prediction errors.

Keywords: Ocean waves(해양파), Abnormality index(이상지수), Multilinear regression(다중선형회귀), Data resampling(자료재추출), Cross validation(교차검증), Rogue waves(거대 파도), Freak waves(기형 파도)

1. 서 론

선박 및 해양, 해안구조물의 설계와 운영, 보호와 안전관리를 위해서 해양파에 대한 연구가 많이 이루어져 왔다. 특히 설계 파고는 선박 및 해양구조물의 설계에서 가장 중요한 인자 중 하나이다. 설계파는 해양에서 발생하는 파도 중 대진폭의 발생과 특성에 영향을 받는다. 해양파에서 일반적으로 예측되는 최대파고 크기를 넘어서는 매우 높은 파고를 가지는 파도를 freak wave 또는 rogue wave라 한다(National Ocean Service[2019]). 북해에 설치된 Draupner platform에서 1995년 1월 1일에 발생한“New Year Wave”라 부르는 파도는 대표적이고 널리 알려진 rogue wave이다(Stansell[2004]).

Rogue wave의 발생 여부는 유의파고와의 파고의 상대비로 표현된 이상지수(AI, abnormality index)를 이용해 다음과 같이 경우로

정의 된다(Kalif *et al.*[2009]).

$$AI = \frac{H}{H_{1/3}} \geq 2$$

여기서 H는 파고이고 H_{1/3}은 유의파고이다. 이상지수의 값은 보통 2~3 사이의 값을 가지며, 가장 큰 값은 North Sea에서 관측된 3.91로 알려져 있다(Divinsky *et al.*[2004]).

Rogue wave에 관한 연구는 그동안 많은 연구자에 의해서 진행됐다. 기존 많은 연구결과는 Kalif *et al.*[2009] 책에 잘 정리되어 있고 참고문헌도 다양하게 포함되어 있다. 기존의 많은 연구는 rogue wave의 발생원인, 발생 확률 및 분포에 대한 것이고, rogue wave 예측모델에 관한 연구는 없는 실정이다. 최근 Draupner platform에서 계측된 파랑자료에 대한 시계열과 주파수영역에서 나타나는 파랑의 주요 15개의 통계 변수에 대한 종합적인 자료가 제시되었고,

[†]Corresponding author: dykim@hongik.ac.kr

각 변수와 이상지수와의 상관성 등 여러 통계 변수의 특성들이 분석되었다(Kim[2019]). 이 논문에서는 rogue wave가 발생한 Draupner platform 파랑관측 자료의 각종 통계 변수들을 바탕으로 다중선형 회귀모델을 사용하여 rogue wave의 판단 기준인 이상지수의 예측 모델을 도출하고, 모델과 데이터의 신뢰성을 검토하고자 한다.

2. 본 론

2.1 파랑 자료

본 논문에서는 노르웨이 근처의 북해에 설치된 Draupner platform에서 1995년 1월에 계속된 파랑 시계열 실측자료를 사용하였다. 각각의 파랑 자료는 20분간 측정이 되었고, 사용된 파랑 자료는 유의 파고가 6 m 이상인 48개의 파랑계측 자료이다. 각 48개의 자료는 ID 번호로 구분을 하며, 총 48개의 자료 중 3개의 자료(ID 4, 17, 46)가 rogue wave를 포함하고 있다. 파랑의 시계열의 자료를 시간 영역과 주파수영역에서 분석하였고, Table 1에는 각종 통계 변수를 표시를 하였다(Kim[2019]). 통계 변수들은 파랑의 시간영역과 주파수영역에서 얻은 파랑의 특성을 나타내는 15개의 변수다. 이 변수들에는 수면의 시계열 자료에서 파형의 분산(σ_η), 비대칭도(γ), 및 첨도(k), 평균파고(H_0), 시계열 평균 영점상향통과주기(T_0), 유의 파고($H_{1/3}$), 파고의 제곱근평균(H_{rms}), 최대파고(H_{max}), 최대파 주기($T_{H_{max}}$)가 포함되어 있다. 주파수영역에서는 유의파고(H_s), 영점상향 통과 주기(T_p), 첨두 주파수(ν), 스펙트럼의 대역폭(S_s), JONSWAP 스펙트럼의 최대치를 조정하는 계수(γ_J)가 포함되어 있다. 이 변수들의 구체적인 값은 부록에 제시하였다.

2.2 다중선형회귀모델

이상지수와 각종 통계변수 중 상관관계가 높은 변수는 큰 순서대로 파형의 첨도(kurtosis), 비대칭도(skewness), 최대파고, 스펙트럼의 대역폭, 유의파첨도 등이 있다. 그리고 주축해석(Principal

Table 1. Description of variables in Draupner wave data

No.	Symbol	Description of Variables
1	σ_η	RMS of wave elevation
2	γ	Skewness of wave elevation
3	k	Kurtosis of wave elevation
4	H_0	Average wave height
5	T_0	Average zero up-crossing period (time domain)
6	$H_{1/3}$	Significant wave height (time domain)
7	H_{rms}	RMS of wave height
8	H_{max}	Maximum wave height
9	$T_{H_{max}}$	Wave period of maximum wave height
10	H_s	Significant wave height (frequency domain)
11	T_z	Average zero up-crossing period (frequency domain)
12	T_p	Peak period of wave spectrum
13	ν	Bandwidth of wave spectrum
14	S_s	Significant steepness
15	γ_J	Peak enhancement factor (JONSWAP spectrum)

Table 2. Coefficients of least square fitting with 15 variables

Coefficient	Variable	Value
β_0	Intercept	-0.55530
β_1	σ_η	2.58880
β_2	γ	-0.02972
β_3	k	0.03551
β_4	H_0	0.32960
β_5	T_0	0.06693
β_6	$H_{1/3}$	-0.06276
β_7	H_{rms}	-0.54246
β_8	H_{max}	0.12894
β_9	$T_{H_{max}}$	0.02731
β_{10}	H_s	-0.76839
β_{11}	T_z	0.19151
β_{12}	T_p	-0.02647
β_{13}	ν	-0.39979
β_{14}	S_s	18.20003
β_{15}	γ_J	-0.02612

coordinate)에서 이상지수 변화량의 고유벡터와의 방향등을 분석하였다(Kim [2019]). 이를 바탕으로 본 논문에서는 주어진 통계 변수로부터 이상지수를 예측하기 위해서, 이상지수가 15개 변수에 선형함수로 가정하고 최소제곱법(least square method)을 사용하여 이상지수 AI 에 대한 다중선형회귀분석을 수행하기로 한다.

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i \tag{1}$$

여기서 y 는 이상지수의 예측치이고, $n = 15$, x_i 는 Table 1에 주어진 15개의 변수이고, β_0 는 회귀직선의 절편 β_n 는 각 변수의 기울기를 나타낸다.

각종 통계계산은 최근 통계계산에 많이 사용하고 있는 R을 사용하였다(The R Foundation[2020]). Table 2에 회귀분석에서 구한 y 축 절편과 기울기의 계수들을 표시하였다. 평균제곱오차(mean square error, MSE)는 다음과 같이 표시된다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - AI_i)^2 \tag{2}$$

다중선형회귀직선과 이상지수 제곱오차합(sum square error, SSE)는 $6.16E-02$ 이고, 평균제곱오차는 $1.28E-03$ 이다. 평균제곱오차와 제곱오차합의 단위는 m^2 이며, 최대파고의 크기와 비교를 한다면 오차의 크기는 크지 않다고 볼수 있다. 다중선형 회귀분석에서는 변수가 15개로 모든 변수에 대해서 표시하면 15차원 함수가 되어서 간단한 그림으로 표시할 수 없어서, 회귀직선에서 얻은 이상지수 예측값과 실제 값과의 관계를 Fig. 1에 표시하여 예측결과를 그래프로 나타내었다. 예측값과 실제값과의 상관관계는 다음과 같이 정의된다.

$$r = \frac{S_{yAI}}{\sqrt{S_{yy}}\sqrt{S_{AIAI}}} \tag{3}$$

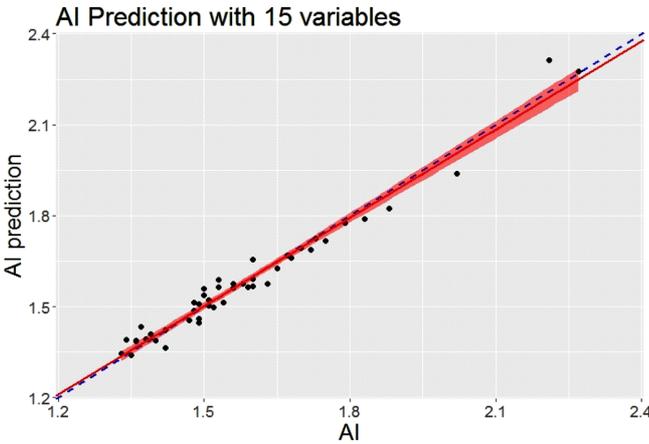


Fig. 1. Least square fitting of 48 data with 15 variables. Shaded zone represents 95% confidence interval of the fitting. (MSE=0.001284, $r=0.984$).

$$S_{yy} = \sum (v_i - \bar{y})^2, S_{AIAI} = \sum (AI_i - \bar{AI})^2,$$

$$S_{yAI} = \sum (v_i - \bar{y})(AI_i - \bar{AI}) \tag{4}$$

이상지수와 예측치의 상관계수 r 은 0.984로 데이터가 회귀직선에 잘 일치하는 것을 볼 수 있다. 점선은 상관관계가 1인 경우를 참고로 표시하였다. 음영으로 표시된 부분은 회귀직선의 95% 신뢰구간을 나타낸다.

Table 3에 데이터 ID 별 이상지수 예측값, 제곱오차(SE)와 내림차순 순위, 전체 제곱오차합(sum square error, SSE) 중 개별 데이터가 차지하는 퍼센트오차를 표시하였다. 이 테이블에는 전체 48개 데이터를 개별 제곱오차가 가장 큰 순서부터 작은 순서로 데이터를 표시하였다. 이 중 ID-4, 17, 46 데이터에는 rogue wave가 포함되어 있고 Fig. 1에서 가장 오른쪽에 있는 3개의 점에서 확인할 수 있다. Rogue wave가 포함된 ID-4 자료가 가장 큰 개별오차를 보이며, ID-17이 두 번째 큰 개별오차를 보이고 있다. 이에 반해 ID-46 데이터는 매우 작은 개별오차를 보인다. ID-4와 ID-17전체 오차의 28%를 차지하고 있는 것을 볼 수 있다.

오차의 절대값을 살펴보면 ID-4에서 실제 이상지수는 2.21인데 예측된 이상지수의 값은 2.31로, 오차가 가장 큰 경우에도 절대오차가 0.10로 오차가 크지 않은 것을 볼 수 있다. ID-17에서는 실제 이상지수는 2.02인데 예측된 값은 1.94이고, ID-46에서는 실제 이상지수는 2.27인데 예측된 값은 2.27이다. 여기서 ID-4가 평가에러가 가장 크나, 이상지수의 값이 가장 큰 2.27인 ID-46의 예측결과가 오히려 회귀직선 결과와 잘 일치하는 것을 볼 수 있다. 이 결과를 종합하면 변수 15개를 이용하여 계산한 회귀직선이 이상지수(AI) 값을 예측하는 데 유용하게 사용할 수 있을 것으로 판단된다.

2.3 교차검증

데이터의 재추출방법(resampling method)은 현대 통계학에서 데이터의 검증과 추가적인 정보를 얻는데 사용되는 분석방법이다

Table 3. Errors of least square fitting of abnormality index with 15 variables (MSE=1.28E-03, SSE=6.16E-02)

ID	AI	AI prediction	SE	SE rank	SE % error
4	2.21	2.31	1.07E-02	1	17.3%
17	2.02	1.94	6.80E-03	2	11.0%
38	1.37	1.43	4.03E-03	3	6.5%
11	1.50	1.56	3.55E-03	4	5.8%
25	1.53	1.59	3.42E-03	5	5.6%
34	1.88	1.82	3.083E-03	6	5.00%
20	1.63	1.57	3.044E-03	7	4.94%
28	1.60	1.65	3.025E-03	8	4.91%
6	1.42	1.37	2.995E-03	9	4.86%
12	1.34	1.39	2.525E-03	10	4.10%
40	1.49	1.45	1.732E-03	11	2.81%
14	1.83	1.79	1.721E-03	12	2.79%
48	1.50	1.54	1.356E-03	13	2.20%
39	1.75	1.72	1.164E-03	14	1.89%
15	1.60	1.57	1.146E-03	15	1.86%
24	1.72	1.69	1.138E-03	16	1.85%
31	1.53	1.56	1.112E-03	17	1.81%
30	1.48	1.51	1.080E-03	18	1.75%
36	1.49	1.46	8.148E-04	19	1.32%
41	1.36	1.39	8.008E-04	20	1.30%
37	1.54	1.51	6.790E-04	21	1.10%
21	1.59	1.57	6.178E-04	22	1.00%
44	1.36	1.38	5.903E-04	23	0.96%
9	1.65	1.63	5.594E-04	24	0.91%
3	1.52	1.50	5.362E-04	25	0.87%
29	1.39	1.41	4.149E-04	26	0.67%
10	1.39	1.41	3.394E-04	27	0.55%
13	1.49	1.51	3.321E-04	28	0.54%
5	1.68	1.66	3.292E-04	29	0.53%
43	1.56	1.58	2.696E-04	30	0.44%
23	1.33	1.35	2.599E-04	31	0.42%
42	1.47	1.45	2.349E-04	32	0.38%
2	1.79	1.78	2.245E-04	33	0.36%
47	1.38	1.39	2.050E-04	34	0.33%
33	1.47	1.46	2.033E-04	35	0.33%
16	1.40	1.39	1.427E-04	36	0.23%
27	1.51	1.52	1.128E-04	37	0.18%
22	1.35	1.34	7.478E-05	38	0.12%
32	1.60	1.59	7.225E-05	39	0.12%
7	1.70	1.69	6.558E-05	40	0.11%
35	1.51	1.50	4.827E-05	41	0.08%
18	1.48	1.49	3.787E-05	42	0.06%
46	2.27	2.27	2.445E-05	43	0.04%
8	1.58	1.58	2.414E-05	44	0.04%
19	1.73	1.73	1.822E-05	45	0.03%
26	1.42	1.42	9.273E-06	46	0.02%
45	1.56	1.56	6.509E-07	47	0.00%
1	1.67	1.67	1.886E-08	48	0.00%

(James et al.[2015])). 모델의 검증을 위해 전체 데이터를 학습세트(training set)와 평가세트(test set)로 나누어 교차검증(cross validation,

CV)을 수행하여 모델과 데이터를 검증하는 방법을 많이 사용한다. 전체 n 개의 데이터 중 $n-1$ 개를 학습데이터로 사용하고, 나머지 1개 교차검증 방법에는 LOOCV(Leave One Out Cross Validation), k -겹 교차검증(k -fold Cross Validation)이 많이 사용된다. LOOCV는 전체 n 개의 데이터를 평가에 사용한다. 이 과정을 전체 n 개 데이터에 차례로 적용을 한다. LOOCV는 편향(bias)이 작은 장점이 있으나 전체

Table 4. Result of LOOCV with 15 variables

ID	Training error (SE)	Training error rank	AI Prediction	Actual AI	Test error (SE)	Test error rank
4	0.0003	48	2.65	2.21	0.1929	1
6	0.0011	46	1.27	1.42	0.0223	2
17	0.0011	47	1.90	2.02	0.0151	3
38	0.0012	45	1.48	1.37	0.0119	4
28	0.0012	42	1.68	1.6	0.0062	5
12	0.0012	40	1.34	1.42	0.0061	6
25	0.0012	44	1.53	1.61	0.0060	7
5	0.0013	31	1.68	1.61	0.0056	8
11	0.0012	43	1.5	1.57	0.0053	9
20	0.0012	41	1.63	1.56	0.0053	10
34	0.0012	39	1.88	1.81	0.0047	11
15	0.0013	35	1.6	1.54	0.0034	12
14	0.0013	38	1.83	1.77	0.0032	13
30	0.0013	34	1.48	1.53	0.0030	14
48	0.0013	36	1.5	1.55	0.0030	15
46	0.0013	12	2.27	2.32	0.0028	16
31	0.0013	33	1.53	1.58	0.0026	17
40	0.0013	37	1.49	1.44	0.0024	18
3	0.0013	29	1.52	1.47	0.0023	19
39	0.0013	32	1.75	1.70	0.0023	20
21	0.0013	27	1.59	1.55	0.0017	21
24	0.0013	30	1.72	1.68	0.0016	22
7	0.0013	15	1.7	1.66	0.0013	23
47	0.0013	22	1.38	1.42	0.0013	24
36	0.0013	28	1.49	1.45	0.0013	25
41	0.0013	26	1.36	1.40	0.0013	26
37	0.0013	25	1.54	1.51	0.0012	27
44	0.0013	24	1.36	1.39	0.0010	28
9	0.0013	23	1.65	1.62	0.0009	29
29	0.0013	21	1.39	1.42	0.0006	30
10	0.0013	20	1.39	1.42	0.0006	31
13	0.0013	19	1.49	1.51	0.0006	32
23	0.0013	18	1.33	1.35	0.0006	33
2	0.0013	16	1.79	1.77	0.0005	34
43	0.0013	17	1.56	1.58	0.0005	35
33	0.0013	13	1.47	1.45	0.0004	36
42	0.0013	14	1.47	1.45	0.0003	37
27	0.0013	10	1.51	1.52	0.0002	38
16	0.0013	11	1.4	1.39	0.0002	39
22	0.0013	9	1.35	1.34	0.0001	40
32	0.0013	8	1.6	1.59	0.0001	41
35	0.0013	7	1.51	1.50	0.0001	42
8	0.0013	5	1.58	1.57	0.0001	43
19	0.0013	4	1.73	1.72	0.0001	44
18	0.0013	6	1.48	1.49	0.0001	45
26	0.0013	3	1.42	1.42	0.0000	46
45	0.0013	2	1.56	1.56	0.0000	47
1	0.0013	1	1.67	1.67	0.0000	48

데이터 수에 해당하는 계산을 수행하기 때문에 계산량이 많다는 단점이 있다. k -겹 교차검증은 전체 데이터를 비슷한 크기를 가진 k 개의 그룹으로 나누고, 이 중 $k-1$ 그룹을 학습에 1개의 그룹을 평가에 사용하는 과정을 k 번 수행을 한다. 일반적으로 데이터의 크기가 큰 경우에는 k 값을 5 또는 10을 많이 사용한다. 이 논문에서는 데이터의 크기가 크지 않아서 k 값으로 3을 사용하였다.

Table 4에 LOOCV의 결과를 평가오차(test error)가 가장 큰 경우부터 작은 순서로 표시를 하였다. 학습오차(training error)의 평균제곱 오차는 1.26E-03, 평가오차(test error)의 평균제곱오차는 6.73E-03이다. 이는 평균제곱근 편차가 각각 0.035와 0.082에 해당한다. 전체적으로 평가오차가 학습오차보다 크게 나타난다. Rogue wave가 포함된 3개의 경우를 살펴보면, ID-4가 평가오차가 가장 크고, ID-17은 2번째, ID-46은 16번째로 평가오차가 큰 것을 볼 수 있다. Rogue wave가 포함되지 않은 ID-6이 두 번째로 평가에러가 크게 나타나고 있다. 이 결과에서도 rogue wave가 포함된 예측자료 ID-4가 평가에러가 가장 크며, 실제 이상지수의 값이 2.32로 가장 큰 이상지수를 가지는 ID-46은 평가에러가 다른 자료에 비해서 크지 않은 것을 볼 수 있다.

Table 5에는 3-fold CV에서 rogue wave가 학습 또는 평가데이터에 포함될 수 있는 8개의 경우를 분류하였다. Case 1은 모든 rogue wave가 학습데이터에 포함되어 있고, 평가데이터에는 없는 경우이다. Case 2~4는 2개의 rogue wave가 학습데이터가 포함되고, 나머지 1개의 rogue wave는 평가데이터에 포함된 경우이다. Case 5~7은 1개의 rogue wave가 학습데이터에 포함되고, 나머지 2개의 rogue wave는 평가데이터에 포함된 경우이다. Case 8은 학습데이터에는 rogue wave가 포함되어 있지 않고, 평가데이터에 3개의 rogue wave가 포함된 경우이다. Case 0은 모든 데이터를 학습데이터로 사용한 경우를 참고로 비교하기 위해서 포함하였다.

Table 6에는 경우별로 학습 및 평가오차가 오차가 큰 상위 3위 경우에 대하여, 이상지수, 이상지수 예측치, 제곱오차 및 오차 순위를 평균제곱오차와 같이 표시하였다. Table 6a에서는 3개의 rogue wave가 모두 학습세트에 포함되어 있으며 ID-46의 경우는 오차 순위가 29번째인 것을 볼 수 있다. Table 6c와 Table 9d에서는 ID-17의 rogue wave는 상위 3위 안에 있고 다른 경우는 상위 3위 안에 포함되어 있지 않지만, 표에 포함하여 표시하였다. Rogue wave가 포함된 3개의 자료 중 ID-4와 ID-17은 대부분 경우 오차가 크게 발생하나, ID-46의 경우는 오차가 평균제곱오차보다 작은 경우도 발생하는 것을 볼 수 있다.

Rogue wave가 발생한 경우 이상지수의 예측결과를 종합적으로 분석하기 위해서, Table 7에는 Table 6에 표시된 총 9가지 경우에서 rogue wave가 포함된 ID-4, 17, 46의 3가지의 경우의 예측된 이상

Table 6. (a) 3-fold CV: Case 1 (Training)

ID	AI	AI prediction	s.e.	error rank
17	2.02	1.92	9.19E-03	1
4	2.21	2.28	4.99E-03	3
46	2.27	2.27	2.31E-05	29
		MSE	1.51E-03	

(b) 3-fold CV: Case 1 (Test)

ID	AI	AI prediction	s.e.	error rank
5	1.68	1.55	1.67E-02	1
12	1.34	1.41	4.89E-03	2
29	1.39	1.44	2.26E-03	3
		MSE	2.10E-03	

(c) 3-fold CV: Case 2 (Training)

ID	AI	AI prediction	s.e.	error rank
17	2.02	1.95	4.83E-03	1
28	1.60	1.67	4.61E-03	2
4	2.21	2.27	3.39E-03	4
		MSE	1.13E-03	

(d) 3-fold CV: Case 2 (Test)

ID	AI	AI prediction	s.e.	error rank
5	1.68	1.45	5.10E-02	1
11	1.50	1.61	1.25E-02	2
46	2.27	2.29	3.84E-04	11
		MSE	5.98E-04	

(e) 3-fold CV: Case 3 (Training)

ID	AI	AI prediction	s.e.	error rank
36	1.49	1.43	3.64E-03	1
4	2.21	2.27	3.25E-03	3
46	2.27	2.27	8.05E-06	28
		MSE	8.53E-04	

(f) 3-fold CV: Case 3 (Test)

ID	AI	AI prediction	s.e.	error rank
38	1.37	1.46	7.75E-03	1
21	1.59	1.51	7.07E-03	2
17	2.02	1.95	5.03E-03	5
		MSE	3.19E-03	

(g) 3-fold CV: Case 4 (Training)

ID	AI	AI prediction	s.e.	error rank
17	2.02	1.97	2.68E-03	1
41	1.36	1.40	1.42E-03	2
46	2.27	2.27	1.53E-05	26
		MSE	3.55E-04	

(h) 3-fold CV: Case 4 (Test)

ID	AI	AI prediction	s.e.	error rank
4	2.21	2.66	2.03E-01	1
30	1.48	1.43	2.29E-03	2
40	1.49	1.45	1.84E-03	3
		MSE	1.32E-02	

Table 5. Cases of Rogue waves in 3-fold CV

Data	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 0
Train	4, 17, 46	4, 17	4, 46	17, 46	4	17	46	-	4, 17, 46
Test	-	46	17	4	17, 46	4, 46	4, 17	4, 17, 46	

Table 6. continued

(i) 3-fold CV: Case 5 (Training)				
ID	AI	AI prediction	s.e.	s.e. rank
4	2.21	2.27	3.81E-03	1
29	1.39	1.44	2.70E-03	2
28	1.6	1.65	2.56E-03	3
		MSE	9.21E-04	
(j) 3-fold CV: Case 5 (Test)				
ID	AI	AI prediction	s.e.	s.e. rank
38	1.37	1.51	1.9EE-02	1
17	2.02	1.92	1.01E-02	4
46	2.27	2.36	8.23E-03	5
		MSE	5.90-03	
(k) 3-fold CV: Case 6 (Training)				
ID	AI	AI prediction	s.e.	error rank
17	2.02	1.98	1.63E-03	1
1	1.67	1.70	8.28E-04	2
31	1.53	1.56	8.01E-04	3
		MSE	2.48E-04	
(l) 3-fold CV: Case 6 (Test)				
ID	AI	AI prediction	s.e.	error rank
4	2.21	2.70	2.37E-01	1
30	1.48	1.42	4.15E-03	2
46	2.27	2.22	2.56E-03	3
		MSE	1.58E-02	
(m) 3-fold CV: Case 7 (Training)				
ID	AI	AI prediction	s.e.	error rank
34	1.88	1.85	7.86E-04	1
9	1.65	1.63	6.08E-04	2
46	2.27	2.27	2.17E-05	27
		MSE	2.03E-04	
(n) 3-fold CV: Case 7 (Test)				
ID	AI	AI prediction	s.e.	error rank
4	2.21	2.63	1.77E-01	1
17	2.02	1.93	7.31E-03	2
6	1.42	1.36	3.87E-03	3
		MSE	1.20E-02	
(o) 3-fold CV: Case 8 (Training)				
ID	AI	AI prediction	s.e.	error rank
14	1.83	1.80	8.25E-04	1
41	1.36	1.39	7.34E-04	2
26	1.42	1.44	5.42E-04	3
		MSE	1.99E-04	
(p) 3-fold CV: Case 8 (Test)				
ID	AI	AI prediction	s.e.	error rank
4	2.21	2.58	1.37E-01	1
17	2.02	1.90	1.40-02	2
46	2.27	2.19	6.19-03	3
		MSE	1.04E-02	

Table 7. Comparison of square errors in 3-fold CV on freak wave occurrence

ID	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 0
4	4.99E-03	3.39E-03	3.25E-03	2.03E-01	3.81E-03	2.37E-01	1.77E-01	1.37E-01	1.07E-02
17	9.19E-03	4.83E-03	5.03E-03	2.68E-03	1.01E-02	1.63E-03	7.31E-03	1.40E-02	6.80E-03
46	2.31E-05	3.84E-04	8.05E-06	1.53E-05	8.23E-03	2.56E-03	2.17E-05	6.19E-03	2.45E-05
SUM	1.42E-02	8.60E-03	8.29E-03	2.06E-01	2.21E-02	2.41E-01	1.84E-01	1.37E-01	1.75E-02

지수의 제곱오차를 표시하였고, 마지막 행에는 각 오차의 합을 표시하였다. 이 Table에서 rogue wave가 학습데이터에 포함된 경우는 음영으로 표시하였다. Table 8에는 Table 7의 결과를 정성적으로 쉽게 보기 위해서 각 ID별 오차의 크기의 case별 순위를 오름차순으로 표시를 하였다. Table 8에서는 오차가 오름 순으로 순위를 표시했다. 따라서 작은 숫자가 오차가 작은 것을 의미한다.

Table 7과 8에서, 오차의 합의 크기는 작은 순서로 case 3, case 2 그리고 case 1이다. 그리고 오차의 합의 크기가 큰 순서로 보면 case 6이 가장 크고 다음은 case 4, case 7인 것을 볼 수 있다. 48개 데이터를 모두 사용한 Case 0이 가장 작은 오차를 보이지 않으며, 평균 정도의 오차를 보이는 것을 볼 수 있다. Case 1의 경우는 모든 rogue wave가 학습데이터에 들어 있으나 이상지수의 예측 오차가 3순위인 것을 볼 수 있다. Rogue wave가 학습데이터에 전혀 포함 되어 있지 않은 case 8이 예측 이상지수가 6순위인 것을 볼 수 있다. Case 3은 개별 ID에 별로 보더라도 ID-4와 ID-46의 오차가 가장 작으며 ID-17에서도 중간 이하의 오차를 가지는 것을 볼 수 있다. 여기서 한 가지 주목할 것은 case 6으로 ID-17만 학습데이터에 들어 있는 경우이다. 이 경우가 가장 큰 오차를 보이며, rogue wave가 학습데이터에 전혀 포함되어 있지 않은 case 8보다 더 큰 오차를 보인다. 교차검증 결과는 주어진 데이터 중 ID-4가 전체 예측모델에 가장 중요한 데이터인 것을 보여준다. 그리고 ID-17에 해당하는 rogue wave는 rogue wave 예측모델에 오차를 크게 할 가능성이 있는 것으로 판단된다.

3. 결 론

본 논문에서는 rogue wave가 포함된 48개의 파랑 계측자료에서 얻은 15개의 통계 변수를 이용하여 이상지수를 예측하는 다중선형 회귀모델에 관한 연구를 수행하였다. 이상지수와 예측치와의 상관 계수는 0.984로 데이터가 회귀직선과 잘 일치하는 것을 볼 수 있었다.

데이터의 재추출(resampling)을 위해서 LOOCV와 k-fold CV를 사용하여 교차검증(Cross Validation)을 수행하였다. LOOCV 결과에서는 ID-4와 ID-17이 큰 오차를 유발하였으나, 이상지수가 가장 큰 ID-46은 작은 오차를 유발하는 것을 볼 수 있었다. 3겹 교차검증 결과를 rogue wave가 학습데이터 또는 평가데이터에 속한 경우에 따른 8가지 경우로 분류해서 이상지수 예측의 오차를 분석하였다. 일반적으로는 데이터가 많을수록 오차가 줄어드는 것을 예상할 수 있으나, rogue wave가 포함된 데이터에 대해서 학습오차와 평가오차를 종합해서 판단했을 때, 모든 데이터를 학습데이터로 사용한 경우가 가장 작은 오차를 가지지 않고 평균적인 오차를 가진다.

Table 8. Error rank of 3-fold CV in ascending order (Small rank less error)

ID	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 0
4	4	2	1	8	3	9	7	6	5
17	7	3	4	2	8	1	6	9	5
46	4	6	1	2	9	7	3	8	5
SUM	3	2	1	8	5	9	7	6	4

ID-4와 46이 학습데이터에 포함된 Case 3에서 이상지수 예측오차가 가장 작은 것을 볼 수 있었으며, ID-17만 학습데이터에 포함된 Case 6이 가장 오차가 크게 나타났다. 교차검증 결과, 데이터 중 ID-4가 예측모델에 중요한 데이터이며, ID-17은 예측모델의 오차를 크게 할 가능성이 있는 데이터로 보인다.

향후 연구로는 데이터에 포함된 각 *rogue wave*의 특성에 관한 비교 연구가 이루어져야 할 것으로 보인다. 또한, 현재 사용한 15개의 통계변수보다 작은 수의 변수를 사용하는 이상 지수 예측모델에 관한 연구도 이루어져야 할 것으로 보인다.

후 기

The author would like to thank Dr. Sverre K. Haver of Statoil for kindly providing the Draupner wave data.

References

- [1] Divinsky, B.V., Levin, B.V., Lopatukhin, L.I., Pelinovsky, E.N., Slyunyaev, A.V., 2004, A freak wave in the Black Sea: observations and simulation. *Doklady Earth Sci* 395A, 438-443.
- [2] James, G, Witten, D., Hastie, T., Tibshirani, R., 2015, *An Introduction to Statistical Learning with Applications in R*(chapter 5 Resample Methods). Springer, New York, 175.
- [3] Kalif, C., Pleinovsky, E., Slunyaev, A., 2009, *Rogue Waves in the Ocean*. Springer Berlin Heidelberg.
- [4] Kim, D.Y., 2019, Statistical Analysis of Draupner Wave Data. *J. Ocean Eng. Technol*, 33(3), 252-258.
- [5] National Ocean Service, 2019, What is a rogue wave?. <https://oceanservice.noaa.gov/facts/roguewaves.html> (Accessed 21 Mar. 2019).
- [6] Stansell, P., 2004, Distributions of extreme wave, crest and trough heights measured in the North Sea. *Ocean Engineering*, 32(8-9), 1015-1036.
- [7] The R Foundation, 2020, The R Project for Statistical Computing. <https://www.r-project.org> (Accessed 21 Mar. 2020).

Received 29 April 2020

Revised 6 July 2020

Accepted 29 July 2020

부 록

Table 1. Draupner wave data statistics (Kim [2019])

ID	σ	γ	k	H_0	T_0	$H_{1/3}$	H_{rms}	H_{max}	$T_{H_{max}}$	H_s	T_z	T_p	v	S_s	γ_J	AI
01	1.922	0.102	2.884	4.80	9.59	7.39	5.33	12.37	10.84	7.69	9.10	11.76	0.404	0.060	2.40	1.67
02	1.943	0.196	3.553	4.65	9.20	7.64	5.34	13.63	11.75	7.76	8.81	13.33	0.434	0.064	1.28	1.79
03	2.790	0.253	3.132	7.10	11.88	10.85	7.85	16.49	15.05	11.14	10.81	15.38	0.450	0.061	1.57	1.52
04	2.983	0.413	4.062	7.12	11.28	11.58	8.20	25.58	12.50	11.92	10.81	15.38	0.475	0.065	1.88	2.21
05	3.020	0.217	3.293	7.57	12.58	12.20	8.55	20.46	14.29	12.04	11.53	18.18	0.461	0.058	0.76	1.68
06	2.849	0.134	2.929	6.79	11.20	11.10	7.70	15.74	17.80	11.37	10.90	15.38	0.488	0.061	1.66	1.42
07	2.991	0.233	2.976	7.83	12.70	11.86	8.70	20.18	15.52	11.96	11.18	15.38	0.481	0.061	1.89	1.70
08	1.730	0.162	2.912	4.33	9.43	6.64	4.81	10.52	9.03	6.88	8.83	11.76	0.414	0.057	1.81	1.58
09	1.546	0.226	2.999	3.70	7.91	5.95	4.20	9.82	10.17	6.15	7.82	10.53	0.393	0.064	2.39	1.65
10	1.666	0.136	2.568	4.34	8.88	6.33	4.71	8.80	8.74	6.62	8.27	11.76	0.384	0.062	1.64	1.39
11	1.554	0.087	2.820	3.83	8.31	5.89	4.27	8.85	11.27	6.19	7.83	10.53	0.428	0.065	2.43	1.50
12	1.522	0.177	3.067	3.69	8.24	6.08	4.21	8.17	9.45	6.06	7.97	10.53	0.383	0.061	2.31	1.34
13	1.783	0.197	2.829	4.47	8.75	6.91	4.94	10.30	10.36	7.12	8.40	11.76	0.403	0.065	1.98	1.49
14	1.788	0.138	3.056	4.51	8.94	6.79	4.98	12.41	10.15	7.14	8.06	11.76	0.480	0.070	1.99	1.83
15	1.965	0.147	2.863	4.92	9.51	7.60	5.48	12.16	9.08	7.82	8.46	11.76	0.490	0.070	2.50	1.60
16	1.832	0.119	2.892	4.43	9.00	7.15	5.01	9.99	10.05	7.31	8.62	11.76	0.428	0.063	2.11	1.40
17	1.512	0.144	3.122	3.68	8.55	5.73	4.13	11.59	11.44	6.01	8.15	11.76	0.435	0.058	1.27	2.02
18	1.586	0.085	2.797	3.89	8.75	6.21	4.36	9.17	10.01	6.32	8.49	11.76	0.424	0.056	1.45	1.48
19	1.576	0.050	3.160	3.96	8.57	6.17	4.41	10.66	9.79	6.28	7.87	9.52	0.399	0.065	3.98	1.73
20	1.571	0.079	2.945	3.76	8.22	6.09	4.27	9.94	8.62	6.24	8.09	10.53	0.421	0.061	2.48	1.63
21	1.738	0.226	3.702	4.08	8.37	6.91	4.74	11.02	10.40	6.95	8.34	10.53	0.420	0.064	3.20	1.59
22	1.982	0.180	2.706	5.04	9.13	7.57	5.54	10.26	10.09	7.91	8.41	11.76	0.451	0.072	2.56	1.35
23	2.000	0.073	2.645	4.93	9.41	7.74	5.50	10.28	11.92	7.99	8.84	10.53	0.417	0.066	4.35	1.33
24	1.865	0.163	2.981	4.79	9.47	7.09	5.26	12.21	9.49	7.43	8.47	11.76	0.452	0.066	2.20	1.72
25	1.696	0.120	2.872	4.15	9.04	6.53	4.62	10.02	11.32	6.76	8.75	11.76	0.414	0.057	1.73	1.53
26	1.539	0.136	2.753	3.75	8.70	5.79	4.17	8.24	9.22	6.13	8.20	11.76	0.441	0.058	1.33	1.42
27	1.552	0.047	2.984	3.59	8.17	5.94	4.11	8.97	10.86	6.18	8.29	11.76	0.438	0.058	1.36	1.51
28	1.694	0.137	3.049	4.22	8.93	6.71	4.72	10.74	11.53	6.75	8.37	10.53	0.435	0.062	2.99	1.60
29	1.854	0.145	2.801	4.64	9.20	7.04	5.14	9.79	11.07	7.38	8.66	11.76	0.438	0.063	2.17	1.39
30	1.884	0.111	2.902	4.81	10.20	7.36	5.32	10.87	11.65	7.52	8.90	13.33	0.504	0.061	1.18	1.48
31	1.566	0.112	2.772	3.94	8.97	5.90	4.30	9.04	10.53	6.23	8.12	11.76	0.480	0.060	1.40	1.53
32	1.554	0.017	2.841	3.77	8.35	5.88	4.24	9.41	10.05	6.17	8.12	10.53	0.403	0.060	2.41	1.60
33	1.539	0.185	3.193	3.68	8.33	5.97	4.18	8.78	8.88	6.15	8.03	10.53	0.413	0.061	2.39	1.47
34	1.552	0.214	3.365	3.76	8.06	5.91	4.23	11.10	10.26	6.17	7.87	10.53	0.413	0.064	2.41	1.88
35	1.641	0.123	3.015	4.02	8.13	6.38	4.52	9.63	9.66	6.56	7.84	10.53	0.406	0.068	2.79	1.51
36	1.565	0.237	2.846	3.85	8.16	5.97	4.28	8.92	8.30	6.24	7.83	10.53	0.422	0.065	2.47	1.49
37	1.514	0.143	2.989	3.66	8.26	5.96	4.13	9.21	9.12	6.04	8.09	11.76	0.427	0.059	1.28	1.54
38	1.557	-0.103	2.613	3.86	8.42	5.96	4.27	8.14	10.24	6.18	8.08	11.76	0.436	0.061	1.37	1.37
39	1.791	0.221	3.085	4.33	8.59	6.85	4.90	12.01	9.78	7.13	8.11	10.53	0.403	0.069	3.39	1.75
40	1.947	0.117	2.761	4.92	9.50	7.61	5.46	11.31	9.71	7.78	8.88	11.76	0.416	0.063	2.46	1.49
41	1.692	0.134	2.884	4.01	8.07	6.41	4.52	8.75	10.19	6.74	8.04	11.76	0.433	0.067	1.72	1.36
42	1.777	0.142	2.928	4.30	8.53	6.87	4.86	10.09	10.03	7.09	8.25	10.53	0.428	0.067	3.34	1.47
43	1.587	0.316	3.210	3.88	8.36	6.14	4.36	9.60	10.33	6.33	7.92	11.76	0.439	0.065	1.46	1.56
44	1.593	0.122	2.685	4.10	8.72	6.24	4.50	8.51	8.10	6.35	8.00	11.76	0.412	0.064	1.47	1.36
45	1.523	0.110	2.826	3.82	8.68	5.93	4.27	9.23	9.39	6.06	8.19	10.53	0.376	0.058	2.31	1.56
46	1.527	0.617	5.286	3.69	7.19	6.10	4.27	13.87	7.98	6.08	6.43	9.52	0.546	0.094	3.71	2.27
47	1.811	0.154	2.778	4.59	8.68	6.96	5.06	9.60	9.18	7.19	8.18	9.52	0.351	0.069	5.31	1.38
48	1.907	0.241	2.995	4.64	8.65	7.37	5.22	11.06	11.87	7.59	8.23	10.53	0.456	0.072	3.89	1.50