MAUSAM, **66**, 3 (July 2015), 617-624

551.509.5

Utilization of categorical and continuous combination for wave model verification

SUBEKTI MUJIASIH and DANANG EKO NURYANTO*

Climatology and Geophysics (BMKG), Jl. Raya Tuban, Kuta - 80362, Badung, Bali, Indonesia

**Research and Development BMKG, Jl. Angkasa I/2 Kemayoran Jakarta Pusat - 10720, Indonesia*

e mail : subekti.mujiasih@bmkg.go.id

सार – BMKG (बी एम के जी) में वर्ष 2004 से पवनतरंग मॉडल से काम लिया जा रहा है। इस मॉडल से प्राप्त हुई सूचना समुद्री उपयोगकर्ताओं के लिए बहुत लाभकारी है और यह मुख्य समुद्री पूर्वानुमान उत्पादन का समर्थन करती है। इस अध्ययन ने वर्ष 2012 में तुंगतामापिकी उपग्रह की तुलना में दैनिक तरंग मॉडल आउटपुट को सत्यापित किया है। गुणवत्ता और जनता के विश्वास को बढ़ाने के लिए इस मॉडल की आवधिक रूप से जाँच और आकलन किया जाना चाहिए। जाँच किया गया प्राचल 24 घंटे आगे (F 24), 48 घंटे आगे (F48) और 72 घंटे आगे (F72) के अग्रकाल के मॉडल से महत्वपूर्ण तरंग ऊँचाई था। आँकड़ों का क्षेत्र इंडोनीशिया के पानी (19 वाटर्स) से 1° X 1° ग्रिड पर लगभग 90°-141° पूर्व, 12°-15° दक्षिण तक रहा। सत्यापन में चार पारंपरिक ऋतुओं, दिसम्बर-जनवरी-फरवरी (डी जे एफ), मार्च-अप्रैल-मई (एम ए एम), जून-जुलाई-अगस्त (जे जे ए) और सितम्बर-अक्तूबर-नवम्बर (एस ओ एन) पर विचार किया गया। इसमें सुस्पष्ट और सतत अभिगम की पद्धति का उपयोग किया गया। सतत अभिगम की पद्धति में रिलेटिव ऑपरेटिंग कैरकटरिस्टिक (आर ओ सी) का उपयोग किया गया कि किस प्रकार तरंग ऊँचाई (मीटर) के तीन निर्णय संबंधी अंतरालों पर आधारित घटना के होने और न होने को पूर्वानुमान कैसे बेहतर ढंग से अलग कर सकता है। आर ओ सी स्कोर से पता चला कि किसी श्रेणी की तरंग का बेहतर पता लगाया जा सकता है। सतत अभिगम से प्रत्येक तरंग श्रेणी के लिए रूट मीन स्कवेयरड एरर (आर एम एस ई) की गणना की गई जिसके लिए अच्छा आर ओ सी स्कोर (≥ 0.7) प्राप्त किया गया। पूर्वानुमान के मान की सटीकता को मापने के लिए आर एम एस ई उपयोगी रहा।

इस अध्ययन से पाँच मुख्य निष्कर्ष निकले। पहला, बहुत से क्षेत्र और ऋतुओं का आर ओ सी स्कोर मान अच्छा नहीं रहा और कुछ का आर ओ सी ए F 24 की अपेक्षा F 48 अथवा F 72 में बेहतर रहा। यह आंशिक रूप से सीमित तुंगतामापी आँकड़ा नमूनों से संबंधित है। दूसरा,गहरे समुद्रों के पूर्वानुमान के लिए पवनतरंग मॉडल अधिक उपयुक्त हैं। तीसरा, पवनतरंग मॉडल ने अग्रकाल F 24, थ्रेशहोल्ड (0-1.25) तथा डी जे एफ ऋतु में अच्छा कार्य निष्पादन किया। चौथा, कुछ समुद्रों का एक से अधिक तरंग श्रेणी में अच्छा आर ओ सी स्कोर रहा जिसका अभिप्राय है कि बहुत सी तरंग ऊँचाईयों का पूर्वानुमान लगाया जा सकता है। पाँचवा, आर एम एस ई स्कोर से पता चला कि पवनतरंग मॉडल पूर्वानुमान करने के लिए एक अच्छा माध्यम है।

ABSTRACT. The Windwaves model has been operating at BMKG since 2004. The model output information is highly desirable by marine users and supports the main marine forecasting product. To increase quality and public trust, this model should be verified and evaluated periodically. This study verified daily wave model output against altimetry satellite during 2012. The examined parameter was significant wave height from the model at lead times of 24 hours ahead (F24), 48 hours ahead (F48) and 72 hours ahead (F72). The data domain extends over Indonesian waters (19 waters) about 90°-141° East, 12°-15° South on a 1° × 1° grid. The verification considered the four conventional seasons, December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON). The method used was a combination of categorical and continuous approaches. The categorical approach used the Relative Operating Characteristic (ROC) to measure how well the forecast can separate events from non-events based on three decision intervals of wave height (meter). The ROC score shows which wave categories can be well-detected. The continuous approach calculated the Root Mean Squared Error (RMSE) for each wave category for which a good ROC score (≥ 0.7) was obtained. RMSE was useful to measure the accuracy of the forecast value.

This study resulted in five main findings. First, many areas and seasons did not have any good ROC score values, and some had ROCA better at F48 or F72 than at F24. This is partly related to the limited altimeter data samples. Second, Windwaves model was more suitable for high seas prediction. Third, Windwaves model performed best in lead time F24, threshold (0-1.25) and DJF season. Fourth, some waters had good ROC scores for more than one wave category, which means that a variety of wave heights are predictable. Fifth, RMSE scores showed that Windwaves model was a good forecasting tool.

Key words – ROC, RMSE, Wave forecasting, Contingency table, Windwaves.

1. Introduction

Indonesia is an archipelago that is dominated by the sea, so the sea inevitably becomes a very important means of transport both for passengers and freight transport. Not only weather on land but also that at sea may be changeable and tends to be extreme. Extreme weather conditions can increase the potential of vessel accidents, especially if the condition of the vessels is inadequate.

For example, KM. Wahai Star passenger ship sank due to a leak in the back of the ship so the water flowed into the engine room. This condition was caused by the collision of the boat when being towed and tied with a short distance (more than 7 m). This crash was caused by weather conditions with high waves 3-4 meters (NTSC, 2007). Another ship accident, KM. Teratai Prima sank due to overload and unbalanced position of charge. As a result, it caused excessive over-draft ship and small inverting moment of ship. So, in the event of bad weather, the ship lost stability and sank (NTSC, 2009). In these circumstances, it is necessary to do in-depth verification research on wave prediction model, subsequently those predictions can be used as a reference to reduce the risk of accidents.

Wave model verification has received more attentions since thirty years ago. With ocean wave models being used in operational mode, appropriate verification of a wave model against observed wind and wave data is necessary and important. The performance of a wave model must be continually assessed to determine its strengths and weaknesses so that it can be improved through adjustment or modifications. It is also necessary to develop sufficient confidence in the model products for operational use.

There are some verification studies using observation data. National Oceanic and Atmospheric Administration/ National Data Buoy Center (NOAA/NDBC) has verified ocean wave ensemble at NCEP using buoy and statistical method such as Brier score (BS), Brier skill score (BSS), reliability diagram, cost-loss analysis, ROC area (Cao *et al*., 2007). Li and Yu (2001) have verified a third-generation wave model (WAM) KMA model against data from buoys using statistical methods such as bias, root-mean-squared error (RMSE), scatter index and correlation coefficient. The World Meteorological Organization (WMO) has recognized the research of wave verification against moored buoy in a technical report by Bidlot and Holt (2006). In this research they studied monthly comparisons with buoy data using basic statistics such as the mean of the difference between models and observations (bias), the root mean square error, the scatter index (SI) defined as the standard deviation of the difference normalized by the mean of the observations, linear correlation coefficient and the mean square slope defined as the ratio of the variance of the model and the variance of the observations (Bidlot and Holt, 2006).

Not only buoy data, satellite data can also be used for verifying. Gusdal *et al*. (2011) have validated the significant wave height from the operational wave model WAM against EnviSat Radar Altimeter (RA-2) and *in situ* observations. The operational wind wave prediction system at KMA was verified against Topex/Poseidon Jason (wave) and Quickscat (wind) data (Park *et al*., 2007). Another study is the verification of the ECMWF Wave Forecasting System, where the analyzed wave height and peak period field were verified against buoy data and first-guess wave height against ERS-1 altimeter data. The forecast skill of the ECMWF wave forecasting system such as error growth in forecast wave height was also studied (Janssen *et al*., 1997).

Recently, the WMO-IOC Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) has developed a verification system involving altimetry satellite data. This system is operated by ECMWF, National Oceanography Centre Southampton, UK (JCOMM, 2013). WMO suggests four methods for verifying the forecast. These are time series of winds and waves; verification spreadsheet of root mean square error, bias and skill score; pie chart for any given threshold value; and target spreadsheet for any given target threshold (WMO/TD-N 850, 1998). The third method used contingency table and pie for displaying.

The contingency table is an input for calculating the relative (or receiver) operating characteristic (ROC) curve (Swets, 1973; Mason, 1982; Harvey *et al*., 1992). Furthermore, the ROC curve is a useful method of representing forecast skill and the area under the curve, *A*, is the most commonly used (and simplest to calculate) and has become known as the ROC score (Mason and Graham, 1999 and 2002). Other literature refers to it as ROC Area (ROCA) (WMO/TD - No. 1485, 2008). A recent study also utilized contingency table to determine wave model performance in terms of the ROC score (Colman *et al*., 2011).

Verification of forecast Indonesia sea waves against altimeter data has been done using RMSE (Kurniawan *et al*., 2013) and contingency table (Mujiasih, 2013). BMKG has developed wave modelling using Wave Watch III for high seas and offshore forecasting, and Simulating Waves Nearshore (SWAN) model for coastal around marine meteorological stations. However, both of these models are still used experimentally. The daily wave forecast uses Windwaves-5.0 as the operational tool. So, Windwaves-5.0 must be verified.

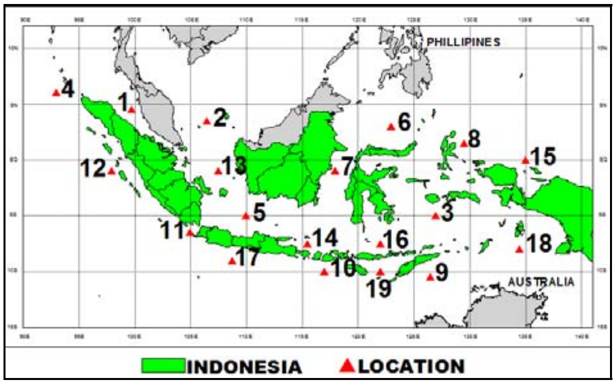


Fig. 1. Indonesian waters map

Following recent studies and WMO documents, this study was conducted to quantitatively verify model wave height against altimeter satellite data using ROCA and RMSE. The aim was to find out which thresholds were predicted well by the Windwaves model for each season and water, allowing precise recommendations for operation to be given. Altimeter satellite data for verification was chosen because it was established and we do not have in-situ wave observation data.

2. Data and methodology

2.1. *Data*

This research included all Indonesia waters from 90°-141° E, 12°-15° S. The location and sample size for the 19 waters in this study are shown in Fig.1 and Table 1. The sample size of data in Table 1 depends on the availability of altimeter data as comparison. The analysis is divided into four seasons: December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON).

The verified data were significant wave height (SWH) of Windwaves 5.0 model output in 2012. Windwaves 5.0 model has been operating since 2004. It was developed by using the existing models especially from Isozaki and Uji (1973). It is deep sea model. The used existing model was MRI-II wave model. This model was used by Japan Meteorological Agency (JMA) for operational forecasting in North-western Pacific (Fujita (1993) in Suratno (1997)). However, some parameters were not considered in this model such as refraction, shoaling and seabed friction near shore (Suratno, 1997).

Model input is 10 meter wind from NOAA NCEP's Global Forecast System (GFS) output. The spatial resolution was 1° × 1° and the temporal resolution was daily, run at 0000 UTC. SWH is defined as the average

TABLE 1

Data sample size for each location

No.	Location	DJF	MAM	JJA	SON
1.	Malacca Strait	54	40	67	57
2.	Natuna Sea	54	50	58	53
3.	Banda Sea	52	32	77	62
4.	Northern Aceh	57	49	65	59
5.	Java Sea	54	35	75	57
6.	Celebes Sea	60	62	72	56
7.	Makassar Strait	51	22	66	52
8.	Halmahera Sea	60	53	77	63
9.	Timor Sea	57	41	80	68
10.	Southern Bali & NTB	54	40	79	67
11.	Sunda Strait	55	32	76	58
12.	Western Sumatera	58	28	62	51
13.	Karimata Strait	46	29	66	53
14.	Bali Sea	55	32	79	61
15.	Northern Papua	57	50	74	60
16.	Flores Sea	52	29	78	63
17.	Southern Java	54	44	79	64
18.	Arafuru Sea	52	29	76	67
19.	Sawu Sea	55	41	79	65

height of the 1/3 highest waves, which roughly approximates to visually observed wave height (WMO-No. 702, 1998). The tested output of Windwaves 5.0 model consisted of 24 hours ahead (F24), 48 hours ahead (F48) and 72 hours ahead (F72). This output was compared against daily wave height from satellite-borne altimeters in 2012.

Altimeter SWH is estimated by analyzing the shape and intensity of the altimeter radar beam reflected from sea surface (radar echo). This data came from merged multi-mission satellite from Jason-2, Envisat, Cryosat and HY-2, available whenever there are at least two missions. Global daily data in NetCDF format can be downloaded by ftp from ftp://ftp.aviso.altimetry.fr/pub/oceano/AVISO/wind-wave/nrt/mswh/merged/. Data are processed using the last 2 days of available data for each satellite and dated from the last of these 2 days. There are two wave height products based on its time availability. There are data in near real time (NRT) and data in delayed time. We used NRT data because it was gridded (1°×1°) making it easier to compare to gridded wave model output.

TABLE 2

List of wave threshold categories

No.	Threshold (meter)
1.	0 < x < 1.25
2.	1.25 < x < 2.0
3.	>2.0

2.2. *Methodology*

In this study, two methods were combined, categorical and continuous. The categorical method was used for analyzing which wave categories were successfully predicted at certain waters for each lead time and each season. Three categories were defined using wave height thresholds (Table 2).

The steps for this method are as follows. First, these categories were verified as events in the contingency table according to table A1 in Appendix A. Types of events were hits, misses, false alarm, correct negatives. The Probability of Detection (POD) score and Probability of False Detection (POFD) score based on these events were used to calculate the Relative Operating Characteristic Area (ROCA) (Appendix A). The area under the ROC curve is frequently used as a score. The range of ROCA value is 0 - 1. ROCA score of 0.5 means no skill system, whereas a ROCA value of 1.0 means system can predict well. The ROCA equal or greater than 0.7 was chosen to show the system predicted well. Finally, calculation using a continuous score, RMSE was carried out. All steps were calculated by combining excel and R software.

The methodology is shown in Fig 2. The result of verification included two outputs. These are ROC and RMSE for each wave category, season and lead time where ROCA $\geq$ 0.7.

3. Results and discussion

The main results are shown in Table 3. These data were derived for each season and lead time where the ROCA score was equal or more than 0.7. The table illustrates some findings.

3.1. *ROCA*

There were some waters where the Windwaves model provided excellent prediction of both the threshold up to 2 meters and also the potential for high waves (more than 2 meters). Windwaves model satisfactorily predicted the threshold until 2 meters for Banda Sea, Java Sea,

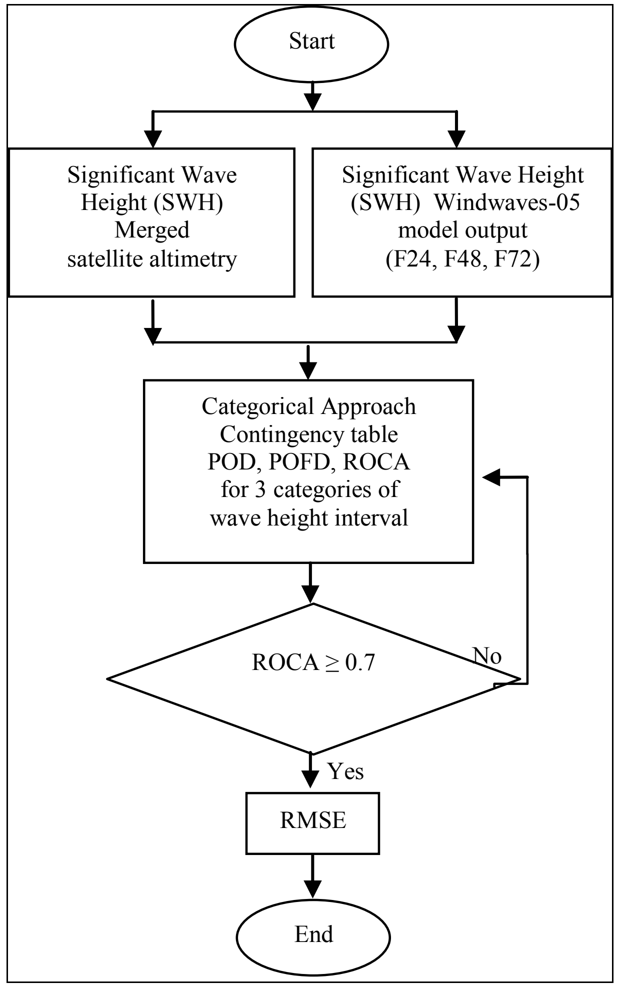


Fig.2. Evaluation methodology

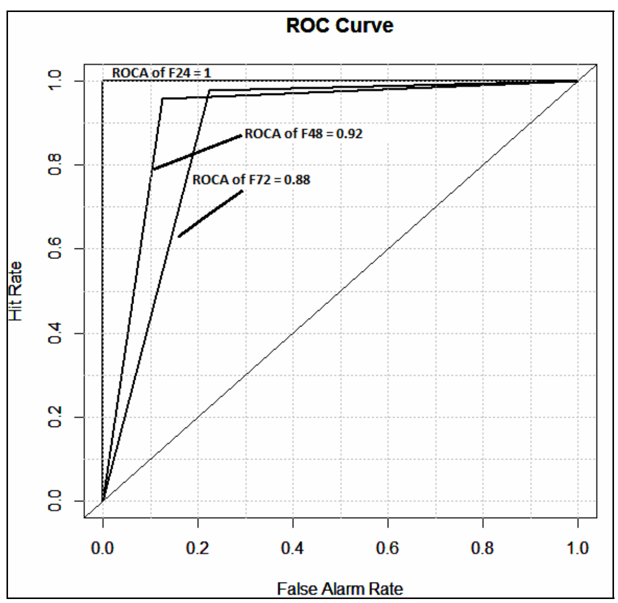


Fig. 3. ROC curve for threshold (0-1.25 m) in DJF-Timor Sea

TABLE 3

ROCA and RMSE for each waters, lead time and season. Only results for ROCA ≥ 0.7 are shown

Location	Season	Number of Samples	Leadtime	Measure	Threshold (meter)		
					0-1.25	1.25-2.0	> 2
Natuna Sea	DJF	54	F24	ROCA	0.77	0.71	0.70
				RMSE		0.77	
Banda Sea	MAM	32	F24	ROCA	0.86	0.78	
				RMSE		0.76	
			F48	ROCA	0.87	0.94	
				RMSE		0.76	
			F72	ROCA	0.85	0.93	
				RMSE		0.71	
Northern Aceh	MAM	49	F24	ROCA	0.76		0.79
				RMSE		0.54	
			F48	ROCA	0.74		0.75
				RMSE		0.57	
			F72	ROCA	0.77		
				RMSE		0.55	
Java Sea	DJF	54	F24	ROCA	0.97		
				RMSE		0.42	
			F48	ROCA	0.96	0.97	
				RMSE		0.45	
			F72	ROCA	0.70	0.70	
				RMSE		0.42	
	MAM	35	F72	ROCA	0.75	0.75	
				RMSE		0.46	
Celebes Sea	DJF	60	F24	ROCA	0.79	0.80	
				RMSE		0.49	
			F48	ROCA	0.75	0.78	
				RMSE		0.60	
			F72	ROCA	0.75	0.78	
				RMSE		0.58	
Halmahera Sea	MAM	53	F24	ROCA	0.78		
				RMSE		0.28	
Timor Sea	DJF	57	F24	ROCA	1.00	1.00	
				RMSE		0.20	
			F48	ROCA	0.92	0.79	
				RMSE		0.29	
			F72	ROCA	0.88	0.79	
				RMSE		0.30	
	MAM	41	F24	ROCA	0.77		
				RMSE		0.58	
Southern Bali & NTB	DJF	54	F24	ROCA			0.70
				RMSE		0.54	
	MAM	40	F24	ROCA	0.86		
				RMSE		0.56	
			F48	ROCA	0.74		
				RMSE		0.76	
Sunda Strait	DJF	55	F24	ROCA	0.73		
				RMSE		0.45	
			F48	ROCA	0.76		
				RMSE		0.43	
Karimata Strait	DJF	46	F24	ROCA	0.95		
				RMSE		0.41	
			F48	ROCA	0.78		
				RMSE		0.58	
			F72	ROCA	0.76		
				RMSE		0.56	
Northern Papua	DJF	57	F24	ROCA	0.83	0.82	
				RMSE		0.24	
Southern Java	DJF	54	F24	ROCA	0.72		
				RMSE		0.56	
	MAM	44	F24	ROCA	0.84		
				RMSE		0.77	
			F48	ROCA	0.70		
				RMSE		0.79	

TABLE 3 (*Contd.*)

Location	Season	Number of Samples	Leadtime	Measure	Threshold (meter)		
					0-1.25	1.25-2.0	> 2
Arafura Sea	DJF	52	F24	ROCA	0.87		
				RMSE		0.44	
			F48	ROCA	0.74		
				RMSE		0.70	
			F72	ROCA	0.74	0.73	
				RMSE		0.62	
	MAM	29	F24	ROCA	0.73		0.72
				RMSE		0.68	
			F48	ROCA	0.77		
				RMSE		0.70	
			F72	ROCA	0.76		
				RMSE		0.81	
	JJA	76	F48	ROCA		0.74	
				RMSE		0.95	
	SON	67	F48	ROCA	0.72	0.77	
				RMSE		0.72	
Sawu Sea	DJF	55	F24	ROCA	0.89	0.89	
				RMSE		0.26	
			F48	ROCA	0.82	0.82	
				RMSE		0.28	
			F72	ROCA	0.78	0.78	
				RMSE		0.28	

Celebes Sea, Timor Sea, Northern Papua and Sawu Sea. Furthermore, Windwaves model also satisfactorily predicted the potential high waves (more than 2 meters) for Natuna Sea, Northern Aceh, Southern Bali & NTB and Arafuru Sea. There were some areas with no ROCA value computed (no altimeter samples) and some less than 0.7 (not shown) and some with perfect score. There were some areas the Windwaves model could not predict well because of having ROCA < 0.6, such as Malacca Strait, Makassar Strait, and Western Sumatera; or having ROCA score between 0.6-0.7 such as Bali Sea and Flores Sea.

Other areas still had good predictions in certain seasons or lead times. For instance, the largest ROCA was 1 for threshold category (0-1.25) and (1.25-2.0) in lead time F24 for Timor Sea (Figs. 3 and 4).

Generally every water had one skilful wave category for each season and lead time. However, some waters may have two or three skilful categories. There were some areas that had two categories (0-1.25) and (1.25-2.0), such as Banda Sea, Java Sea, Celebes Sea, Timor Sea, Northern Papua and Sawu Sea. There were some areas that had two categories (0-1.25) and (>2) such as Northern Aceh and Arafuru Sea. As a matter of fact, Natuna Sea was the only water which had full categories ((0-1.25), (1.25-2.0), (>2 meters) in DJF season. It means Natuna Sea has wide variation of skilful wave height categories for DJF season.

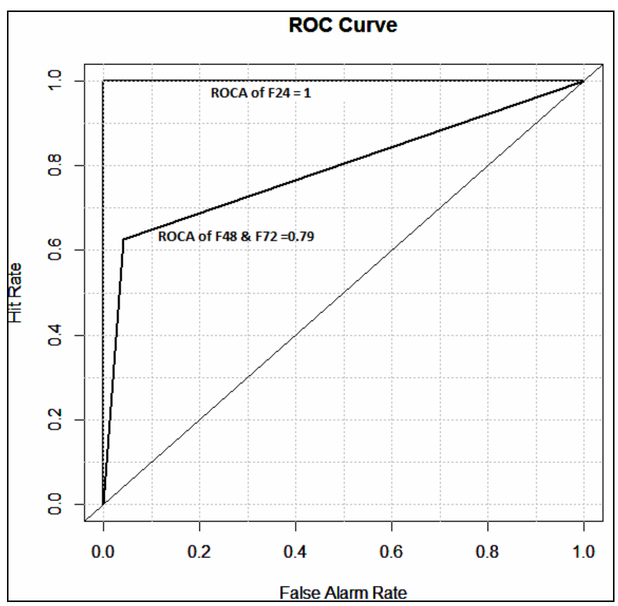


Fig. 4. ROC curve for threshold (1.25-2 m) in DJF-Timor Sea

In general the wave category could be predicted well just for one or two seasons. Concerning skilful categories, some areas had it for one season, some had it for two seasons and some had it for all seasons. The areas which had one predictable season were Natuna Sea, Banda Sea, Northern Aceh, Celebes Sea, Halmahera Sea, Sunda

Strait, Karimata Strait, Northern Papua and Sawu Sea. The areas which had two predictable seasons were Java Sea, Timor Sea, Southern Bali & NTB and Southern Java. Arafuru Sea was the only water that had all seasons predictable. The system usually could predict well in lead time F24. There were some times when the Windwaves model could not predict well in lead time F24, such as in MAM for Java Sea and in JJA and SON for Arafuru Sea. On the whole the best wave category which could be predicted well was threshold (0-1.25). It is shown by the biggest ROCA from each lead time (Table 3).

Overall the seasons the Windwaves model could predict well were in DJF and MAM. There were 11 areas which had good performance in DJF including Natuna Sea, Java Sea, Celebes Sea, Timor Sea, Southern Bali & NTB, Sunda Strait, Karimata Strait, Northern Papua, Southern Java, Arafuru Sea and Sawu Sea. Then, there were eight areas which had good performance in MAM including Banda Sea, Northern Aceh, Java Sea, Halmahera Sea, Timor Sea, Southern Bali & NTB, Southern Java and Arafuru Sea. The Windwaves model could not predict well in SON and JJA seasons except in the Arafuru Sea.

Finally, the system had good capability to predict outer waters (nearby Indian Ocean, Pacific Ocean, or South China Sea) such as Natuna Sea, Northern Aceh, Celebes Sea, Halmahera Sea, Timor Sea, Southern Bali & NTB, Sunda Strait, Northern Papua, Southern Java, Arafuru Sea and Sawu Sea. It proves that Windwaves model is a skilful deep sea model as mentioned earlier.

3.2. *RMSE*

This study came up with two results for RMSE. First, the average of RMSE for all results was below 1 m. This value is still good and accepted. Second, the RMSE score for inner waters of Indonesia generally had lower values than for outer waters, below about 0.5 for Java Sea, Karimata Strait, Sunda Strait and Sawu Sea.

3. Conclusions

Based on the analyses above, the following conclusions can be made:

(*i*) There were many areas and seasons which had no good ROCA value, and some wave categories with no ROCA value at all. This is partly due to the limited altimeter samples.

(*ii*) Windwaves model was more suitable for high seas prediction.

(*iii*) Windwaves model predicts best at lead time F24, threshold (0-1.25) and DJF season.

(*iv*) Some waters have more than one skilful wave category, indicating a wide variation of wave prediction.

(*v*) RMSE score showed Windwaves model was still good for forecasting tool.

In further studies this model will be verified against in-situ observation from rigs or platforms. Subsequently, it will be compared with the verification result from other models, such as WaveWatch III and SWAN.

References

Bidlot, J. R. and Holt, M. W., 2006, “Verification of operational global and regional wave forecasting systems against measurements from Moored buoys”, JCOMM Technical Report No. 30, WMO/TD-NO. 1333.

Cao, D., Chen, H. S. and Tolman, H. L., 2007, “Verification of Ocean Wave Ensemble Forecast at NCEP, NOAA/NWS/NCEP/EMC/MMAB Technical Note Nr. 261. Available at http://polar.ncep.noaa.gov/mmab/papers/ tn261/.

Colman, Andrew W., Palin, Erika J., Sanderson, Michael G., and Harrison, Robert T., 2011, “The Potential for Seasonal Forecasting of Winter Wave Heights in the Northern North Sea”, *Wea. Forecasting*, **26**, 1067-1074.

Gusdal, Yvonne., Carrasco, Ana., Furevik, Birgitte R., and Sætra, Øyvind., 2011, “Validation of the Operational Wave Model WAM”, Norwegian Meteorological Institute Met. No, Report No. 14/2011 Oceanography, ISSN: 1503-8025 Oslo, December 8, 2011.

Harvey et Al., 1992, “The Application of signal detection theory to weather forecasting behavior”, *Mon. Weather Rev.,* 120,863-883

Isozaki, I. and Uji, T., 1973, “Numerical prediction of ocean waves”, *Paper in Met. and Geophys*., **24**, 207-232.

Janssen, P. A. E. M., Hansen, B. and Bidlot J. R., 1997, “Verification of the ECMWF Wave Forecasting System against Buoy and Altimeter Data”, *Wea. Forecasting*, **12**, 763-784.

JCOMM, 2013, http://www.jcomm.info/index.php?option=com_content &view=article&id=131&Itemid=37, accessed on 30th July, 2013

Jolliffe, I. T. and Stephenson, D. B., 2012, “Forecast Verification: A Practitioner's Guide in Atmospheric Science”, John Wiley & Sons.

Kurniawan, R., Permana, D. S., Suratno, Habibie, M. N., 2013, “Verification of Windwaves-5.0 output against satellite altimetry (in Indonesian)”, *Journal of Meteorology and Geophysics*, **14**, 149-158.

Li, J. and Yu, Z., 2001, “Verification of the WAM4 Wave Model, Proceedings The Sixth Workshop on Ocean Models for the APEC Region”, September 11-14, 2001, Beijing, P. R. China.

Mason, I.,1982,"A Model for Assessment of weather forecasts", *Aust.Meteorol.Mag.*,30,291-303.

Mason, S. J. and Graham, N. E., 1999, "Conditional probabilities, relative operating characteristics and relative operating levels", *Wea. Forecasting*, **14**, 713-725.

Mason, S. J. and Graham, N. E., 2002, "Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation", *Q. J. R. Meteorol. Soc*., **128**, 2145-2166.

Mujiasih, S., 2013, "Assessment of Altimetry Satellite Data For Windwaves Model Verification", Asian Pacific Coast of Remote Sensing 13-19 October, 2013, Bali.

NTSC, 2007, "The sinking of KM", Wahai Star Pulau Tiga Nusa Telu, Ambon July 10, 2007. Maritime Accident Investigation Report of National Transportation Safety Committee (NTSC), NTSC-07-07-05-03.

NTSC, 2009, "The sinking of KM", Teratai Prima Tanjung Batu Roro Waterway, South Sulawesi January 11, 2009, Maritime Accident Investigation Report of National Transportation Safety Committee (NTSC), NTSC-09-01-01-03.

Park, S., Lee, Da-un., and Seo, Jang-won., 2007, "Operational wind wave prediction system at KMA", JCOMM Scientific and Technical Symposium on Storm Surge, 2-6 October 2007, COEX, Seoul, Korea, JCOMM-TR-44.

Suratno, 1997, "Numerical model of Surface Wave Forecasting for Indonesia waters", Thesis Graduate, University of Indonesia.

Swets, J. A., 1973, The relative operating characteristic in psychology. Science, 182, 990–1000

Wilks, D. S., 2006, "Statistical Methods in the Atmospheric Sciences", Academic Press.

WMO/TD-N 850, 1998, "Handbook of Offshore Forecasting Services", Report No.36, Marine Meteorology and Related Oceanographic Activities.

WMO-N0.702, 1998, "Guide to Wave Analysis and Forecasting", 2nd edition, World Meteorological Organization.

WMO/TD - No. 1485, 2008, "Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models", WWRP 2009 - 1, Revision 2, October 2008.

Appendix A

Verification metrics

TABLE A1

Contingency Table

		(O) Observed		
		Yes	No	Total
	Yes	hits	False alarm	Forecast Yes
(F) Forecast	No	Misses	Correct Negatives	Forecast No
	Total	Observed yes	Observed No	Total

Hit rate/Probability of Detection (POD)

$$\mathrm{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}} \tag{A.1}$$

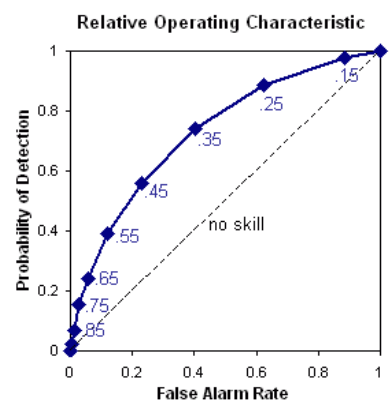
This score can be explained as what fraction of the observed "yes" events were correctly forecast. The range is 0 to 1. The perfect score is 1. The characteristics are sensitive to hits but ignoring false alarms. This score is very sensitive to the climatological frequency of the event and good for rare events (Wilks, 2006; Jolliffe and Stephenson, 2012).

False Alarm Rate/Probability of False Detection (POFD)

$$\mathrm{POFD} = \frac{\text{false alarms}}{\text{correct negatives} + \text{false alarms}} \tag{A.2}$$

This score can be explained as what fraction of the observed "no" events were incorrectly forecast as "yes". The range is 0 to 1. The perfect score is 0. The characteristics are sensitive to false alarms but ignoring misses. (Wilks, 2006; Jolliffe and Stephenson, 2012).

Relative Operating Characteristic (ROC)



The relative operating characteristic is a representation of the skill of a forecast system in which the hit rate and the false-alarm rate are compared (Swets, 1973). The area under the curve, *A*, is the most commonly used (and simplest to calculate) and has become known as the ROC score (Mason and Graham, 1999; Wilks, 2006). Other literature called this the *ROC Area (ROCA)* (WMO/TD - No. 1485, 2008). Ranges from 0 to 1, 0.5 indicates no skill. Perfect score is 1.

Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is a metric for continuous variables which is widely used. This score is frequently used to examine the error magnitude. Smaller RMSE indicates the system has better performance. The equation for RMSE is given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(F_i - O_i\right)^2}$$

where F_i is the forecast value, O_i is the observed value, and N is the number of samples. Range is 0 to ∞. Perfect score is 0. Characteristics are simple and familiar. Measures "average" error, weighted according to the square of the error. RMSE value does not indicate the direction of the deviations (Wilks, 2006).