REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
08-02-2004		

4. TITLE AND SUBTITLE
Data Assimilation Modeling of the Barotropic Tides in the Korea/Tsushima Strait

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0601153N

6. AUTHOR(S)
Jeffrey Book, Pavel Pistek, Henry Perkins, Keith R. Thompson, William J. Teague, Gregg A. Jacobs, Moon-Sik Suk, Kyung-Il Chang, Jae-Chul Lee and Byung Ho Choi

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
73-7607-C3-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
Oceanography Division
Stennis Space Center, MS 39529-5004

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/JA/7330-03-17

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5660

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
During 1999-2000, 13 bottom mounted acoustic Doppler current profilers (ADCPs) and 12 wave/tide gauges were deployed along two lines across the Korea/Tsushima Strait, providing long-term measurements of currents and bottom pressure. Tidally analyzed velocity and pressure data from the moorings are used in conjunction with other moored ADCPs, coastal tide gauge measurements, and altimeter measurements in a linear barotropic data assimilation model. The model fits the vertically averaged data to the linear shallow water equations in a least-squares sense by only adjusting the incoming gravity waves along the boundaries. Model predictions are made for the O_1, P_1, K_1, $\cdot_2$, N_2, M_2, S_2, and K_2 tides. An extensive analysis of the accuracy of the M_2 surface-height predictions suggests that for broad regions near the mooring lines and in the Jeju Strait the amplitude prediction errors are less than 0.5 cm. Elsewhere, the analysis suggests that errors range from 1 to 4 cm with the exception of small regions where the tides are not well determined by the dataset. The errors in the model predictions are primarily caused by bias error in the models physics, numerics, and/or parameterization as opposed to random errors in the observational data. In the model predictions, the highest ranges in sea level height occur for tidal constituents M_2, S_2, K_1, O_1, and N_2, with the highest magnitudes of tidal velocities occurring for M_2, K_1, S_2, and O_1. The tides exhibit a complex structure in which diurnal constituents have higher currents relative to their sea level height ranges than semi-diurnal constituents.

15. SUBJECT TERMS
Korea/Tsushima Strait, barotropic tide, data assimilation, error analysis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Jeffrey W. Book
Unclassified	Unclassified	Unclassified	UL	17	19b. TELEPHONE NUMBER *(Include area code)* 228-688-5251

Journal of Oceanography, Vol. 60, pp. 977 to 993, 2004

Data Assimilation Modeling of the Barotropic Tides in the Korea/Tsushima Strait

JEFFREY W. BOOK[1]*, PAVEL PISTEK[1], HENRY PERKINS[1], KEITH R. THOMPSON[2], WILLIAM J. TEAGUE[1], GREGG A. JACOBS[1], MOON-SIK SUK[3], KYUNG-IL CHANG[3], JAE-CHUL LEE[4] and BYUNG HO CHOI[5]

[1]*Naval Research Laboratory, Stennis Space Center, MS 39529-5004, U.S.A.*
[2]*Dalhousie University, Halifax, NS B3H 4J1, Canada*
[3]*Marine Environmental and Climate Change Laboratory, Korea Ocean Research and Development Institute, Ansan P.O. Box 29, Seoul 425-600, Korea*
[4]*Pukyong National University, 599-1 Daeyondong, Namgu, Pusan 608-737, Korea*
[5]*Sung Kyun Kwan University, Suwon 440-746, Korea*

(Received 8 September 2003; in revised form 6 February 2004; accepted 8 February 2004)

During 1999–2000, 13 bottom mounted acoustic Doppler current profilers (ADCPs) and 12 wave/tide gauges were deployed along two lines across the Korea/Tsushima Strait, providing long-term measurements of currents and bottom pressure. Tidally analyzed velocity and pressure data from the moorings are used in conjunction with other moored ADCPs, coastal tide gauge measurements, and altimeter measurements in a linear barotropic data assimilation model. The model fits the vertically averaged data to the linear shallow water equations in a least-squares sense by only adjusting the incoming gravity waves along the boundaries. Model predictions are made for the O_1, P_1, K_1, μ_2, N_2, M_2, S_2, and K_2 tides. An extensive analysis of the accuracy of the M_2 surface-height predictions suggests that for broad regions near the mooring lines and in the Jeju Strait the amplitude prediction errors are less than 0.5 cm. Elsewhere, the analysis suggests that errors range from 1 to 4 cm with the exception of small regions where the tides are not well determined by the dataset. The errors in the model predictions are primarily caused by bias error in the model's physics, numerics, and/or parameterization as opposed to random errors in the observational data. In the model predictions, the highest ranges in sea level height occur for tidal constituents M_2, S_2, K_1, O_1, and N_2, with the highest magnitudes of tidal velocities occurring for M_2, K_1, S_2, and O_1. The tides exhibit a complex structure in which diurnal constituents have higher currents relative to their sea level height ranges than semi-diurnal constituents.

Keywords:
· Korea/Tsushima Strait,
· barotropic tide,
· data assimilation,
· error analysis.

1. Introduction

The Korea/Tsushima Strait joins the East China Sea and the Yellow Sea (both areas of strong tides) to the Japan/East Sea[§] (an area of weak tides). Ogura (1933) made the first comprehensive study of the tides in the strait using coastal station data and provided a basic understanding of the tidal heights, tidal dynamics, and amphidromic points. Since this study, many others have investigated the tides in the area, making use of both numerical models (Fang and Yang, 1988; Choi *et al.*, 1989; Kang *et al.*, 1991; Kantha *et al.*, 1996) and observations (Odamaki, 1989; Kang *et al.*, 1995). These studies all agree in a qualitative sense and show the structure of the tidal heights in the strait. Semidiurnal tidal height amplitudes peak along the south coast of Korea, with a general trend from high values in the East China Sea to low values in the Japan/East Sea; amphidromic points for these constituents occur just north of the strait in the Japan/East Sea proper. Semidiurnal tidal height phases remain constant through much of the strait, but increase rapidly from west to east across the northern boundary. Diurnal tidal height amplitudes range from high values in the East China Sea to low values in the Japan/East Sea; amphidromic points for these constituents occur in the northwestern corner of the

* Corresponding author. E-mail: book@nrlssc.navy.mil

§ The Editor-in-Chief does not recommend the usage of the term "East Sea" in place of "Japan Sea".

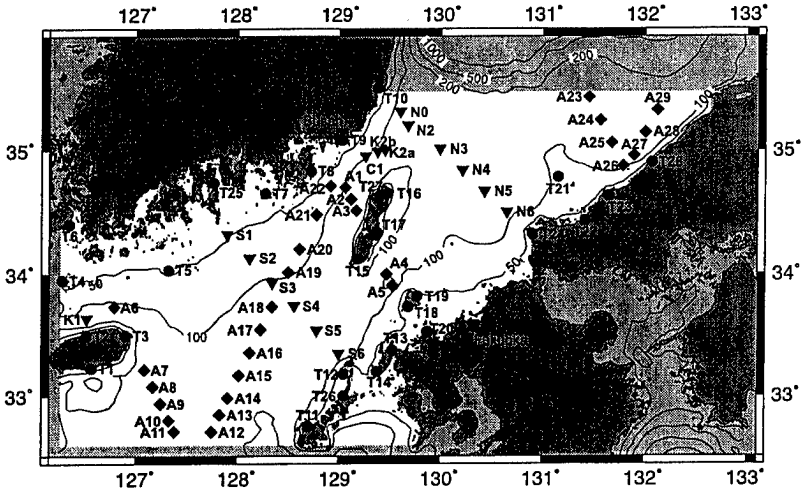

Fig. 1. Locations of observational moorings (triangles), coastal tide stations (circles), and TOPEX/POSEIDON altimeter measurements (diamonds). Bathymetric contours for the 50, 100, 200, 500, and 1000 m isobaths are calculated from the SKKU dataset and drawn as thin lines. The cities of Busan (Pusan) and Fukuoka (squares) are shown for reference. Dark shading indicates land areas. Light shading indicates ocean areas not considered by the data assimilation model. Jeju (Cheju) Island is the large island to the south of Korea in the far western portion of the map. Jeju Strait is the strait between Jeju Island and Korea. Tsushima Island is the large pair of islands in the center of the map. The Gotou Islands are the cluster of islands in the southern portion of the map to the west of the main Japanese island of Kyushu. Honshu is the main Japanese island in the northeast portion of the map.

strait, close to the Korean coast. Diurnal tidal height phases increase along an arc from the south coast of Korea to the northwest coast of Honshu. Fang and Yang (1988), Odamaki (1989), and Teague *et al.* (2001) show that tidal currents are strong in many portions of the Korea/Tsushima Strait.

Further work is needed to obtain sufficiently accurate tidal height and current predictions for the Korea/Tsushima Strait. Many numerical models have poor resolution in the Korea/Tsushima Strait and therefore cannot resolve complex tidal structures or represent the interaction of the tides with complex topography. Other studies use finer resolution models, but limit the number of tidal constituents that they consider. These regional modeling studies often force their models at the open ocean boundaries with values from historical co-tidal charts or larger scale models, both of which are subject to error. Observational studies have been limited, away from the coast, to short time series of tidal current data in which some tidal constituents are not resolvable. Moreover, such studies do not predict tidal currents away from the observation stations.

Odamaki (1989) edited the tidal charts of Ogura (1933) based on a large observational dataset that included data from many short-duration current meters. Odamaki (1989) used the new charts produced in this comprehensive study to examine the tidal dynamics in the Korea/Tsushima Strait. Recently, Teague *et al.* (2001) have shown that the Korea/Tsushima tidal charts need further adjustment. They analyzed data from an initial deployment of 11 moorings, each equipped with an acoustic Doppler current profiler (ADCP) and a pressure gauge, along two lines crossing the strait. These moorings gathered current and pressure data and were deployed as a part of the U.S. Naval Research Laboratory (NRL) "Linkages of Asian Marginal Seas" (LINKS) program. This program will be described in more detail in Section 2. Tidal heights were obtained by harmonic analysis of the pressure time-series and then compared to the corrected chart values from Odamaki (1989). Chart amplitudes along the southern line of moorings are about 10% high for the semidiurnal tides and 10% low for the diurnal tides. A direct comparison between the tidal current coefficients obtained from a LINKS mooring and a nearby historical tidal current observation show a 40% difference in M_2 amplitude and a 100% difference in O_1 amplitude. The most likely cause for these differences is that the current observations from Odamaki (1989) were of too short duration to accurately resolve the various tidal frequencies and therefore he used various assumptions to account for unresolved constituents.

This paper uses the tidal characterizations at 13 stations obtained during the LINKS study and other data with a barotropic data assimilation model developed by Griffin and Thompson (1996). The model produces predictions of tidal currents and tidal heights by using the

Table 1. Calculated tidal amplitudes (amp.) and phases (pha.) in cm and degrees for the mooring pressure time series (Fig. 1). Nodal corrections were applied to these amplitudes and phases. The phases are the Greenwich phase lags according to the convention given by Foreman (1977). The estimated amplitude errors (err.) are in cm and were calculated as described in Subsection 3.2.

	N0	N2	N3	N4	N5	N6	S1	S2	S3	S4	S5	S6
O_1 amp.	2.21	2.34	3.67	5.14	6.91	9.12	14.28	13.34	13.27	13.57	14.78	16.10
O_1 err.	0.10	0.06	0.07	0.08	0.11	0.10	0.16	0.15	0.11	0.12	0.13	0.14
O_1 pha.	194.2	171.1	150.8	144.1	144.0	144.9	34.7	44.1	53.4	61.3	68.7	74.7
P_1 amp.	0.58	0.26	0.81	1.35	1.97	2.66	6.23	5.80	5.64	5.65	5.89	6.32
P_1 err.	0.10	0.06	0.07	0.08	0.11	0.10	0.17	0.15	0.11	0.12	0.14	0.15
P_1 pha.	259.5	197.6	159.6	156.7	149.5	162.9	52.0	59.3	67.4	74.7	79.8	88.7
K_1 amp.	1.34	0.98	2.60	4.33	6.38	8.43	19.88	18.41	17.86	17.79	18.60	19.95
K_1 err.	0.10	0.06	0.07	0.08	0.11	0.10	0.17	0.15	0.11	0.12	0.14	0.15
K_1 pha.	266.7	206.5	165.6	160.5	162.3	166.7	58.5	65.8	73.3	80.2	87.5	93.4
μ_2 amp.	1.02	0.92	1.05	1.16	1.10	1.39	4.62	4.40	4.19	4.08	3.40	4.04
μ_2 err.	0.09	0.07	0.07	0.08	0.08	0.09	0.16	0.14	0.11	0.14	0.13	0.14
μ_2 pha.	334.4	348.8	1.8	10.7	21.9	24.0	349.2	350.6	351.6	353.0	357.0	356.5
N_2 amp.	3.14	3.16	3.45	3.77	3.96	4.59	17.84	16.53	15.62	15.05	14.24	14.79
N_2 err.	0.09	0.07	0.07	0.08	0.08	0.09	0.16	0.14	0.11	0.14	0.13	0.15
N_2 pha.	323.2	336.3	352.4	2.5	9.7	19.9	352.5	352.1	351.5	351.0	350.2	350.4
M_2 amp.	16.28	17.08	18.28	19.86	21.32	23.43	86.16	80.09	76.09	73.87	72.11	72.85
M_2 err.	0.09	0.07	0.07	0.08	0.08	0.09	0.16	0.14	0.11	0.14	0.13	0.15
M_2 pha.	333.0	344.0	357.8	7.2	14.4	23.9	1.8	0.9	0.1	359.0	357.4	358.0
S_2 amp.	9.22	9.25	9.96	10.70	11.00	12.29	40.14	37.52	35.81	34.79	33.04	34.27
S_2 err.	0.09	0.07	0.08	0.08	0.08	0.09	0.16	0.15	0.11	0.14	0.13	0.15
S_2 pha.	4.0	12.7	23.7	31.2	36.9	45.0	26.1	25.2	24.2	23.3	21.5	22.2
K_2 amp.	2.69	2.48	2.75	2.93	2.84	3.42	11.36	10.61	10.15	9.79	8.73	9.71
K_2 err.	0.09	0.07	0.08	0.08	0.08	0.09	0.16	0.15	0.11	0.14	0.13	0.15
K_2 pha.	353.0	3.2	14.6	22.0	27.6	35.8	17.7	16.4	15.1	14.2	12.9	12.6

data to infer optimal open ocean boundary conditions for the model. This paper also analyzes the accuracy of the predictions and estimates the accuracy of the tide model.

2. Tidal Observations

Figure 1 shows the locations of the mooring data (triangles), coastal tide station data (circles), and satellite altimeter data (diamonds) used in this work. Thirteen of the sixteen moorings (N0–N6, S1–S6, C1) were part of the LINKS array. Most LINKS sites were instrumented from May 1999 through March 2000 with an ADCP and a wave/tide gauge, housed in a trawl-resistant bottom mount. The ADCPs measured currents throughout the water column and the wave/tide gauges measured pressure at the bottom. Due to mooring failure, instrument failure, or limited numbers of instruments, no current or pressure data are available from N0 and C1 for May through October, no current or pressure data are available from S5 for October through March, and no pressure data are available from C1 and N5 for October through March. Further details of this mooring effort can be found in Teague *et al.* (2002). Two additional bottom-mounted ADCPs were deployed during the time of the LINKS moorings by the Korea Ocean Research and Development Institute (KORDI) and Pukyong National University. The KORDI mooring, K1, was deployed in the Jeju (Cheju) Strait from March through December of 1999. The Pukyong National University mooring, K2a and K2b, was deployed in the West Channel of the Korea/Tsushima Strait from June through December of 1999. This mooring was struck by fishing gear and moved at least three times. The data from the KORDI mooring and the data from the first two positions (June through September, October through November) of the Pukyong National University mooring are used in this study. The data from the third position of the latter mooring are not used because the ADCP failed shortly after this last move.

A least-squares harmonic tidal analysis was performed on all current and pressure data measured at the mooring sites. Prior to this analysis, the current data were vertically averaged to obtain estimates of the barotropic currents. The tidal analysis was done using the Foreman (1977) tidal analysis software adapted to MATLAB by Lentz, Beardsley, and Pawlowicz. The analysis by Teague *et al.* (2001) of the initial deployment (May through October) of LINKS moorings found that the eight largest

Table 2. Calculated tidal major semi-axes (semi-maj.) and minor semi-axes (semi-min.) amplitudes, ellipse inclination angles (inc.), and phases (pha.) in cm/s and degrees for the velocity time series from the north and south mooring lines (Fig. 1). Nodal corrections were applied to these amplitudes and phases. The phases are the Greenwich phase lags according to the convention given by Foreman (1978). Negative minor semi-axes amplitude values indicate clockwise rotation of the tidal current ellipse. The estimated amplitude errors (err.) are in cm/s and were calculated as described in Subsection 3.2.

	N0	N2	N3	N4	N5	N6	S1	S2	S3	S4	S5	S6
O_1 semi-maj.	8.43	8.23	9.48	10.07	10.15	11.20	9.79	9.59	9.43	9.97	9.31	11.43
O_1 semi-min.	−2.46	−3.41	−3.51	−3.99	−2.58	−1.02	1.68	−0.76	−1.40	−2.06	−1.36	−1.55
O_1 err.	0.29	0.16	0.12	0.14	0.17	0.15	0.14	0.16	0.19	0.15	0.16	0.13
O_1 inc.	44.9	37.8	37.0	30.7	36.0	40.6	26.8	29.3	33.8	32.2	46.6	41.1
O_1 pha.	100.0	109.4	118.1	130.7	142.1	148.4	132.7	134.7	131.9	126.9	118.3	114.2
P_1 semi-maj.	3.44	3.38	3.72	3.68	3.71	4.07	3.69	3.56	3.65	3.87	3.49	4.25
P_1 semi-min.	−1.32	−1.11	−1.51	−1.47	−1.20	−0.30	0.40	−0.40	−0.36	−0.72	−0.60	−0.53
P_1 err.	0.30	0.16	0.12	0.14	0.17	0.15	0.15	0.16	0.20	0.16	0.16	0.13
P_1 inc.	50.6	39.5	43.6	36.5	41.3	43.7	23.6	29.3	33.3	35.3	51.2	40.0
P_1 pha.	112.6	133.8	143.5	155.8	166.7	172.8	154.9	161.0	155.0	151.1	142.5	138.7
K_1 semi-maj.	11.74	11.24	12.10	11.93	11.47	12.01	11.38	11.18	10.80	11.88	10.62	12.68
K_1 semi-min.	−2.75	−4.83	−5.10	−5.34	−3.12	−1.53	1.44	−0.78	−1.03	−2.57	−1.49	−1.71
K_1 err.	0.30	0.16	0.12	0.14	0.17	0.15	0.15	0.16	0.20	0.16	0.16	0.13
K_1 inc.	47.9	43.2	43.5	38.9	40.5	44.5	24.6	28.7	32.9	30.9	44.0	40.7
K_1 pha.	123.7	137.7	145.9	158.6	171.9	179.6	162.8	164.0	160.9	156.4	145.3	144.0
μ_2 semi-maj.	1.84	1.43	1.07	1.07	0.82	0.94	1.07	0.83	0.72	0.69	0.63	0.92
μ_2 semi-min.	−0.24	−0.23	−0.05	−0.04	0.01	0.01	0.03	0.07	−0.03	−0.16	−0.23	−0.09
μ_2 err.	0.20	0.11	0.08	0.10	0.11	0.11	0.10	0.09	0.10	0.09	0.10	0.09
μ_2 inc.	52.3	56.1	58.7	55.0	60.4	47.8	178.7	178.9	0.5	9.9	35.3	18.8
μ_2 pha.	58.0	67.3	70.8	73.0	81.3	80.4	259.8	266.0	81.8	73.3	42.0	48.2
N_2 semi-maj.	5.84	4.74	4.15	3.86	3.72	3.97	4.16	3.61	2.96	2.83	1.94	2.86
N_2 semi-min.	−0.80	−0.92	−0.78	−0.59	−0.51	−0.34	0.51	−0.41	−0.81	−1.65	−1.26	−1.12
N_2 err.	0.20	0.11	0.08	0.10	0.11	0.11	0.10	0.09	0.10	0.09	0.10	0.09
N_2 inc.	56.9	53.2	57.5	54.9	53.3	52.4	174.0	2.9	3.0	0.7	41.7	27.3
N_2 pha.	64.9	66.3	80.1	76.4	83.2	82.4	293.0	115.4	110.9	101.5	54.2	64.9
M_2 semi-maj.	26.85	23.82	20.49	18.81	17.58	18.96	19.10	15.62	12.73	12.04	9.44	12.11
M_2 semi-min.	−2.97	−4.52	−3.51	−3.11	−1.92	−0.92	3.46	−1.35	−4.58	−7.52	−6.43	−6.16
M_2 err.	0.20	0.11	0.08	0.10	0.11	0.11	0.10	0.09	0.10	0.09	0.10	0.09
M_2 inc.	57.8	56.6	58.5	53.9	54.2	51.6	176.8	0.9	0.2	178.8	44.6	34.1
M_2 pha.	76.1	76.3	87.4	86.6	90.9	92.2	310.0	131.8	124.5	295.4	63.8	70.6
S_2 semi-maj.	12.62	10.64	8.88	8.33	7.73	8.26	7.84	6.23	5.05	4.67	3.49	4.86
S_2 semi-min.	−1.34	−1.95	−1.43	−1.42	−0.68	−0.40	1.62	−0.18	−1.68	−2.83	−2.55	−2.33
S_2 err.	0.21	0.11	0.08	0.10	0.11	0.11	0.10	0.10	0.10	0.09	0.10	0.09
S_2 inc.	58.9	56.3	58.6	53.2	53.0	52.4	170.8	177.7	175.2	178.0	40.8	33.9
S_2 pha.	101.8	101.0	110.1	111.0	114.1	115.7	332.8	334.6	329.9	316.7	85.6	92.0
K_2 semi-maj.	4.01	2.36	2.33	2.23	2.29	2.53	2.22	1.85	1.56	1.40	1.28	1.33
K_2 semi-min.	−0.83	−0.76	−0.81	−0.58	−0.32	−0.24	0.43	−0.18	−0.59	−0.71	−0.83	−0.56
K_2 err.	0.21	0.11	0.08	0.10	0.11	0.11	0.10	0.10	0.10	0.09	0.10	0.09
K_2 inc.	56.6	56.5	64.2	50.5	57.8	56.3	178.9	1.2	2.1	166.3	44.5	37.4
K_2 pha.	96.5	72.9	99.7	104.1	108.7	110.7	331.6	151.8	136.0	321.2	82.2	79.7

tidal constituents in the pressure data were O_1, P_1, K_1, μ_2, N_2, M_2, S_2, and K_2. Excluding low frequency and shallow water constituents, these are also the eight largest constituents in the histogram of mean tidal amplitudes calculated from 192 tide gauges in the East China Sea, the Yellow Sea, and the Korea/Tsushima Strait by Lefèvre *et al.* (2000). Therefore, these eight constituents were used for the analysis of the mooring data in this study. The resulting tide parameters are presented in Tables 1–3. Records from K2a and K2b were too short to resolve the constituent pairs of P_1/K_1 and S_2/K_2, and therefore P_1 and K_2 were not included in the analysis of those particular time series.

Both the ADCPs and wave/tide gauges in this study were configured so that each measurement was an average over the sampling period as opposed to instantaneous samples. The Foreman (1977) tidal analysis package is meant to analyze instantaneous samples and therefore corrections are required to prevent bias in the calculated amplitudes and phases. Using averages instead of discrete

Table 3. As Table 2, but for the moorings in the Jeju Strait and in the West Channel.

	K1	C1	K2a	K2b
O_1 semi-maj.	7.77	13.86	10.44	13.00
O_1 semi-min.	−0.18	1.00	−0.76	0.28
O_1 err.	0.14	0.31	0.23	0.34
O_1 inc.	5.0	46.9	38.6	47.3
O_1 pha.	200.4	121.9	132.7	120.1
P_1 semi-maj.	2.86	5.01	—	—
P_1 semi-min.	−0.32	0.50	—	—
P_1 err.	0.14	0.32	—	—
P_1 inc.	2.8	45.5	—	—
P_1 pha.	218.2	146.7	—	—
K_1 semi-maj.	9.39	17.17	14.26	16.02
K_1 semi-min.	−0.34	0.49	−0.97	0.84
K_1 err.	0.14	0.32	1.24	3.16
K_1 inc.	6.6	46.2	46.0	46.0
K_1 pha.	229.9	146.8	150.8	133.0
μ_2 semi-maj.	0.93	2.35	1.24	1.92
μ_2 semi-min.	0.06	0.05	0.03	−0.17
μ_2 err.	0.14	0.24	0.20	0.29
μ_2 inc.	3.4	56.6	61.0	48.8
μ_2 pha.	170.1	69.2	123.3	56.0
N_2 semi-maj.	8.36	6.80	4.84	7.14
N_2 semi-min.	0.06	0.25	−0.17	−0.56
N_2 err.	0.14	0.24	0.20	0.29
N_2 inc.	3.7	45.4	47.2	43.6
N_2 pha.	203.5	75.5	84.0	94.0
M_2 semi-maj.	42.13	31.50	26.10	31.93
M_2 semi-min.	−0.67	0.34	−0.73	−1.24
M_2 err.	0.14	0.24	0.20	0.29
M_2 inc.	3.0	45.5	46.6	46.0
M_2 pha.	216.2	86.6	89.7	88.8
S_2 semi-maj.	16.45	14.42	10.59	14.98
S_2 semi-min.	−0.39	−0.03	−0.55	−0.43
S_2 err.	0.14	0.25	0.96	2.31
S_2 inc.	3.4	46.9	46.9	46.8
S_2 pha.	249.1	110.8	115.2	96.7
K_2 semi-maj.	4.89	4.78	—	—
K_2 semi-min.	−0.19	0.08	—	—
K_2 err.	0.14	0.25	—	—
K_2 inc.	4.0	45.8	—	—
K_2 pha.	245.3	101.9	—	—

values causes the tidal analysis software to slightly underestimate tidal amplitudes. This bias is

$$\frac{2}{\omega\Delta t}\sin\left(\frac{\omega\Delta t}{2}\right), \quad (1)$$

where ω is the angular frequency of the tidal constituent and Δt is the measurement sampling period. The amplitudes calculated by the tidal analysis software were divided by this correction factor, which is very nearly equal to 1 for these data because Δt values were either 15 or 30 minutes. The ADCPs and wave/tide gauges assign the start time of the sampling period to each data point. To obtain correct phase information from the analysis, each data point was instead assigned the time of the middle of the sampling period, i.e. adjusted by $\Delta t/2$.

Differences in amplitudes between Tables 1 and 2 and the tables of Teague *et al.* (2001) are small and are mainly because this study uses data from the entire duration of the mooring deployments as opposed to the duration of the initial deployment. However, differences in phases are not small because the time adjustment of $\Delta t/2$ required by the interval averaging was not applied in the analysis of Teague *et al.* (2001), leading to errors in their tidal phase values.

Further tide data for this study are provided by a tidal analysis of TOPEX/POSEIDON (T/P) data and from coastal tide station data (Fig. 1). The T/P data consist of tidal height coefficients calculated from selected points on tracks 107, 64, 18, 56, and 102 using the method described in Teague *et al.* (2000). This calculation estimates tidal coefficients using aliased frequencies because the T/P altimeter does not sample quickly enough to measure the actual tidal frequencies. Selection was done to reduce the T/P data spatial density compared to that of the mooring data and varied between using every other point along the track to using every fourth point along the track. The coastal tide station data were provided as constituent coefficients and are from the International Hydrographic Office (IHO) dataset. Of the numerous tidal stations located in the Korea/Tsushima Strait, 28 were selected for this study based on duration of their measurements, lack of complexity in the coastal area where the instrument was placed, and overall agreement with other data records.

The data assimilation model (Section 3) requires time series data. Therefore, the Foreman (1977) tidal analysis software was used to synthesize time series from the coefficients obtained by the analysis of the mooring data and from the IHO coefficients. The time series thus generated are for a period of 41 days, starting in May of 1999, with values every 30 minutes. Time series in this same form were synthesized from the T/P coefficients using other software in a manner consistent with their coefficient format. Some of the IHO stations did not provide data for the eight constituents considered in this study, and therefore two time series datasets were made. The first dataset was generated at all data sites, but using only the O_1, K_1, M_2, and S_2 coefficients. The second dataset was generated from all eight constituents but excluded those sites without a complete set of constituent data, i.e. T3, T7, T8, T13, T14, T16, T17, T20, T23, T24, K2a, and K2b were excluded. These two datasets provide the forcing for runs of the data assimilation model.

3. Barotropic Data Assimilation Model

3.1 Model dynamics

A numerical model was used to assimilate the available data and provide predictions for tidal currents and heights throughout the Korea/Tsushima Strait. The model uses the linearized barotropic shallow water equations;

$$\frac{\partial u}{\partial t} = -g\frac{\partial \eta}{\partial x} + fv - \frac{\lambda}{h}u, \tag{2}$$

$$\frac{\partial v}{\partial t} = -g\frac{\partial \eta}{\partial y} - fu - \frac{\lambda}{h}v, \tag{3}$$

and

$$\frac{\partial \eta}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} = 0, \tag{4}$$

where u and v are the eastward (x) and northward (y) velocities, respectively, η is the surface height, t is time, g is the gravitational acceleration, f is the Coriolis parameter, h is the depth, and λ is a frictional parameter.

The open ocean boundary conditions provide the control variables for the model as now described. The boundary condition that is used is a standard Reid and Bodine (1968) scheme. This scheme can be interpreted in the following way for the model. Consider the left open ocean boundary. At this boundary, the normal flow can be considered to be a superposition of an incoming (externally forced) gravity wave and an outgoing (radiating) gravity wave. The velocity at the boundary, u_L, is given by

$$u_L = u_i + u_o, \tag{5}$$

where u_i is the velocity of the incoming wave and u_o is the velocity of the outgoing wave. If α is an appropriate admittance coefficient for these gravity waves, then, at the left boundary, u_o equals negative α multiplied by the surface height of the outgoing wave, η_o. In turn, η_o is equal to the total surface height minus the surface height due to the incoming wave, η_i. Using these expressions, the left boundary condition becomes

$$u_L = u_i - \alpha(\eta_L - \eta_i) \tag{6}$$

or

$$u_L = \frac{F}{h} - \alpha\eta_L, \tag{7}$$

where the terms associated with the incoming wave have been combined as one forcing term, F/h. The admittance coefficient for the model was set equal to $\sqrt{g/h}$, corresponding to the admittance of a plane shallow water wave over such a depth.

The open ocean boundaries are divided into a series of structure functions. Along each structure function, the term F from Eq. (7) is allowed to vary in time but is forced to remain constant along its segment of the boundary. A total number of 28 structure functions are used, spaced along the boundaries such that the bathymetry is relatively constant for each structure function. They varied in length from 5 to 75 km. The model is started from a state of rest and is then forced by the 28 time series of F along the boundaries.

3.2 Cost function and weights

A cost function is used to measure how well a particular model run successfully predicts the observed data set. We define the cost function to be

$$J = \Sigma_{j,k}\left(m_{j,k} - p_{j,k}\right)W_j\left(m_{j,k} - p_{j,k}\right), \tag{8}$$

where j is an index of sites and types of data to be used, k is the temporal index of the data, m is the collection of all the u, v, and η measurements, p is the corresponding model predictions for these measurements produced by linear interpolations from the model space and time grids, and W is a weighting function. The best fit to the observed data set is the model state that minimizes the cost function, J, assuming that model-data differences are normally distributed with zero means and W is determined by the inverse of the error variance of the measurements (Thacker, 1987). In practice, the model predictions, p, are known to contain bias from model physics, numerics, and parameter errors (see Subsection 4.4).

The cost function is implicitly a function of the F terms from Eq. (7). Choosing values of F at each structure function for each control time completely determines the values of p and thus J. Optimal values of F are found by minimizing J, subject to the strong constraints of Eqs. (2)–(4) bounded by Eq. (7). The method of Lagrange multipliers is used for this purpose, producing a set of adjoint equations that the multipliers must satisfy (Thacker, 1987). The system of equations comprised of the discrete forms of Eqs. (2)–(4) and (7), and their adjoints are solved, yielding the gradients of Eq. (8) with respect to each structure function value (F). This technique is completely described in Griffin and Thompson (1996) and summarized in Hallock *et al.* (2003). The minimum of the cost function is then found using this gradient in a conjugate gradient method. Values of F are started from zero and then adjusted to the values that minimize

the cost function through this process. Convergence was achieved by 25 iterations of the method.

To optimize the formulation of Eq. (8), the weighting of the measurements must be optimized. Assuming that measurement errors are uncorrelated between measurement locations and are stationary in time, the optimal weighting for Eq. (8) is

$$W_j = \frac{c}{\sigma_{m_j}^2} \tag{9}$$

where $\sigma_{m_j}^2$ is the error variance of the measurements and c is a constant term for all values of j. Unless it is too small or too large for the computer's numerics to handle accurately, the value of c does not affect the solution, i.e. only the relative weights of the measurements are important.

Estimates of $\sigma_{m_j}^2$, and thus W_j, were made using the mooring data and information about the IHO and T/P data. The error in measuring individual velocity or pressure values is negligible compared to the error in separating the tidal oscillations from non-tidal oscillations in the data time series. Background ocean continuum energy causes uncertainty in the least-squares harmonic analysis. To estimate the level of this energy, spectra were calculated for each mooring time series, high energy density peaks were set to the median energy density value of the 0.5–2.5 day^{-1} frequency range, and a straight line was fitted to the adjusted spectral values through this range. Estimates of the background continuum were taken as values of the straight line at K_1 and M_2 frequencies. Errors in the least-squares harmonic analysis for the diurnal and semi-diurnal constituents were calculated in the standard way using a variance matrix (calculated from the time vectors and tidal frequencies) and a random white-noise level that would produce the same energy density as these calculated K_1 or M_2 values. Tables 1–3 present the estimated amplitude errors for the mooring data using this method.

For the T/P time series, the variance matrix was obtained by simulating a seven-year time vector sampled at the T/P repeat frequency. The error calculations were then done using this variance matrix and random white-noise at the average level of all the pressure mooring time-series. The estimated constituent amplitude errors are 1.1–1.2 cm. Potential errors due to non-tidal energy at the aliased frequencies were assumed to be small and not considered. The periods of these aliased frequencies in days are 45.71 for O_1, 88.89 for P_1, 173.19 for K_1, 20.31 for μ_2, 49.53 for N_2, 62.11 for M_2, 58.74 for S_2, and 86.60 for K_2.

For the IHO coefficients, time vectors were simulated with 1 hour sampling for the duration given in the IHO database. The error calculations were done considering all the diurnal and semi-diurnal constituents present in the IHO record to form the variance matrix and using random white-noise at the average level of all the pressure mooring time-series. At some short duration stations, the main eight constituents plus additional constituents were included in the analysis despite violation of the Rayleigh criteria by constituents close in frequency. The standard error analysis using the variance matrix accounts for the magnification of background error caused by this inclusion. Alternatively, at other short duration stations, some of the main eight constituents were not included in the analysis. At these sites, the non-analyzed or missing constituent of the Rayleigh pair will interfere with the analyzed constituent, leading to errors proportional to the amplitude of the missing constituent. For the error calculations, the unknown amplitude of the missing constituent was taken as the model prediction from the eight-constituent run. This same procedure was done to account for errors at moorings K2a and K2b caused by the missing P_1 and K_2 constituents.

Errors caused by limited precision in the reported IHO coefficients were also estimated using the model predictions from the eight-constituent run at the data sites. At several sites the precision error was larger than the analysis error and therefore was used in place of the analysis error. The final estimated error for the IHO coefficients varied widely from site to site with a maximum at station T7 of 6.00 cm for S_2 and a minimum at station T28 of 0.06 cm for each constituent. The former error is large because of the lack of inclusion of K_2 in the analysis of this 3 month record and the latter error is small because the station at Hamada was of 10 year duration and had high precision in the reported results.

Values for $\sigma_{m_j}^2$ were calculated from the estimated amplitude errors of the constituents and then used in Eq. (9) to calculate separate W_j for the four and eight constituent runs. Numerous preliminary runs were done, changing various weights to test the model's sensitivity to them. For most of the weight adjustments, changes to the final solution were small. However, in some regions of the model, such as south of Jeju Island, changing the weights can significantly alter model predictions.

3.3 Model bathymetry and friction parameter

The model domain is an Arakawa C grid with a 2.5 km grid spacing. The bathymetry field, h, is defined every 2.5 km, while u, v, and η are defined every 5 km. Several different bathymetries were used in preliminary runs. The model was first run at a lower resolution (5 km) using both the publicly available U.S. Navy DBDB-V bathymetry database (Naval Oceanographic Office, 1997) and a bathymetry field (SKKU dataset) produced from a 1-minute resolution data set available from the Labora-

tory for Coastal and Ocean Dynamics Studies, Sung Kyun Kwan University (Choi, 1999). The value of the final cost function was 52% higher for the DBDB-V dataset, and therefore of these two datasets, only the SKKU dataset was used at the 2.5 km resolution runs. The high resolution runs with the SKKU dataset produced good results but some position-dependent differences remained between the model predictions and the data, seemingly related to bathymetry. The SKKU dataset does not have deep enough values for the deep trough that exists in the West Channel between Tsushima and Korea. For example, observed depths for the two positions of K2 were 228 and 156 m respectively, but the SKKU dataset gives depths of only 161 and 122 m. The SKKU dataset also appears to be inaccurate near mooring site S1. Further, the SKKU dataset does not include several of the larger islands along the south coast of Korea.

Although use of the SKKU dataset produces good results that are significantly better than using the DBDB-V dataset, this study used a third bathymetric database for the final model runs. This final bathymetry database was formed by gridding (Smith and Wessel, 1990) bathymetry soundings obtained from the National Imagery and Mapping Agency (NIMA) Digital Nautical Charts (DNC®). Land masking was done using the World Vector Shoreline land mapping from the Generic Mapping Tool (Wessel and Smith, 1998) at high resolution and fitting the results to the 2.5 km grid of this study. The resulting final bathymetric database contains finer resolution features than either the DBDB-V or the SKKU datasets, better represents the true depths in the West Channel trough, and produces lower minimized cost function values than runs made with the SKKU dataset.

The value of λ in Eqs. (2) and (3) was varied between 2.3×10^{-4} and 2.3×10^{-2} m/s for different model runs. The value of the cost function decreases and then increases as the friction parameter is lowered, and a value of $\lambda = 6.8 \times 10^{-4}$ m/s was chosen for the final model runs because this parameter value produced the lowest cost function value. The functional relationship between cost function value and frictional parameter is nearly flat near this point (for example, $\lambda = 4.6 \times 10^{-4}$ m/s produces nearly the same cost function value) but rises steeply for λ values greater than 1.1×10^{-3} m/s. The magnitude of the slope also increases for λ values lower than 4.6×10^{-4} m/s but this trend could not be fully evaluated because the numerics of the model failed for friction parameters lower than 2.3×10^{-4} m/s. Kang *et al.* (1995) estimated linearized bottom friction coefficients in different seasons for the Korea/Tsushima Strait. Their early spring value (converted to the form used in this paper) of 1.3×10^{-3} m/s and their summer value of 1.8×10^{-3} m/s are both higher than our optimized value.

In summary, this data assimilation technique attempts to match observed values of u, v, and η by adjusting the forcing along the model boundary. The model is constrained to follow the discrete representation of the dynamical equations at all the internal grid points. Further details of this model and data assimilation technique are presented in Griffin and Thompson (1996).

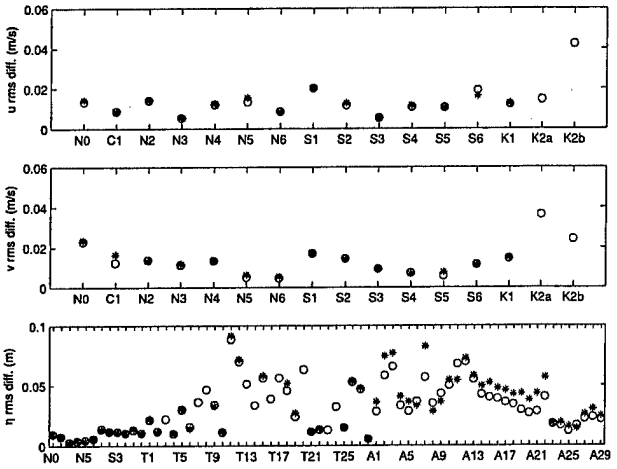

Fig. 2. Rms differences between the observation time series and the model prediction time series. Open circles indicate rms differences from the four-tidal constituent run. Stars indicate rms differences from the eight-constituent model run. Stars are not plotted for sites that lack data for all eight constituents. See Fig. 1 for locations of the data sites.

4. Results and Error Analysis

4.1 Model/data comparisons

How well does the data assimilation model predict the observations? This question could be answered by examining the minimized cost function (Eq. (8)). However, it is more revealing to examine the results of a root mean square (rms) difference calculation, as shown in Fig. 2. This figure shows the rms of the observation time series minus the model predicted time series at all data sites for both the four-constituent run (circles) and the eight-constituent run (stars). The rms differences for both runs are very similar.

Rms differences between the measurements and predictions of either component of velocity are below 4.3 cm/s at all sites and below 2.4 cm/s at all sites except K2a and K2b. Although located close to each other, the tidal parameters measured at C1, K2a, and K2b all had different values (Table 3). The differences between the measurements at C1 and K2a are exceptionally large. Measurements made at K2a and K2b have much higher K_1 and S_2 errors than measurements at other sites, but the

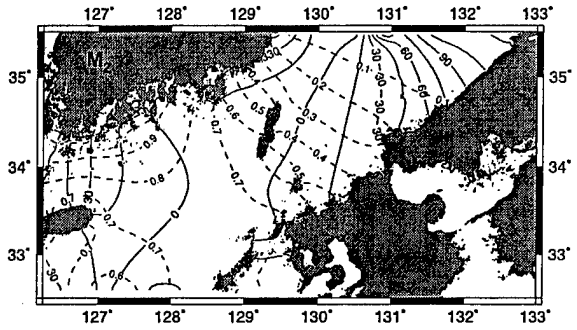

Fig. 3. Model predicted co-tidal (solid) and co-range (dashed) lines for the M_2 tide. The co-range lines are in meters with increments of 0.1 m. The co-tidal lines are in degrees with increments of 15 degrees. The Greenwich phase convention used in this figure is described by Foreman (1977).

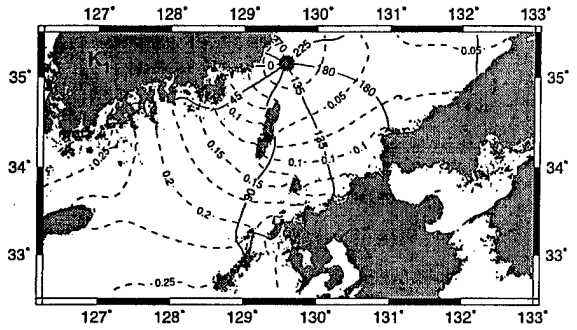

Fig. 5. As Fig. 3, but for the K_1 tide. The co-range lines are in increments of 0.025 m. The co-tidal lines are in increments of 45 degrees. The gray circle indicates the position of an amphidromic point

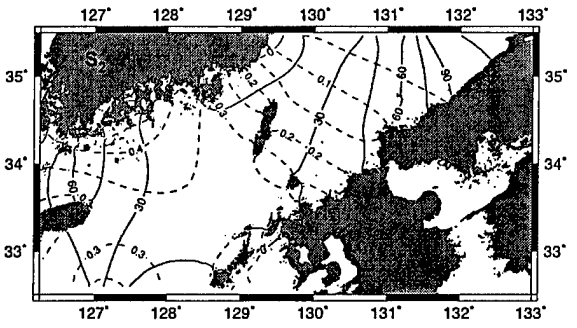

Fig. 4. As Fig. 3, but for the S_2 tide. The co-range lines are in increments of 0.05 m. The co-tidal lines are in increments of 15 degrees.

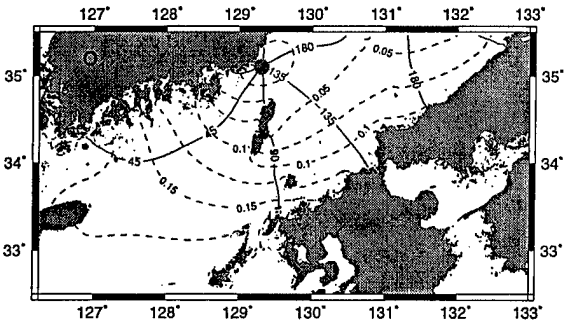

Fig. 6. As Fig. 3, but for the O_1 tide. The co-range lines are in increments of 0.025 m. The co-tidal lines are in increments of 45 degrees. The gray circle indicates the position of an amphidromic point.

differences between the measured values at C1 and K2a exceed these errors. These measured small-scale differences may be caused by bathymetric effects at the deep trough where these moorings were deployed. Alternatively, it is also possible that an unknown source of error has contaminated the measurements at K2a. Regardless, because the estimated errors for C1 are lower than those for K2a and K2b, the data assimilation model gave preference to the C1 data.

The average value of the velocity component rms differences is 1.2 cm/s for the eight constituent run and 1.4 cm/s for the four constituent run. The average values of the rms differences for surface height are 0.9 cm for the LINKS mooring sites, 3.2 cm for the coastal tide stations, and 4.3 cm for the T/P sites using the predictions from the eight-constituent run. For the four-constituent run, the values are 0.9 cm for the LINKS mooring sites, 3.4 cm for the coastal tide stations, and 3.7 cm for the T/P sites. Highest differences occur along the west coast of the Gotou Islands, off the southeast coast of Jeju Island, in the West Channel, and along the edge of the southern model boundary.

Differences between observation and modeled time series are small compared to the total tidal fluctuations in the Korea/Tsushima Strait. The average standard deviations for the observed eight-constituent tidal time series are 13.1 cm/s for the velocity time series, 38.9 cm for the surface height time series at the LINKS moorings, 49.1 cm for the surface height time series at the tidal stations, and 42.5 cm for the surface height time series at the T/P sites. The normalized rms difference for the velocity is less than 0.15 for all time series except for the v time series at S1, S2, K1, and K2a and the four-constituent u time series at S6 and K2b. Variances of the observed v time series at S1, S2, and K1 are small. The normalized rms difference for surface height is less than 0.15 for all time series except for time series at sites A2–A3 and at sites A23–A29. Variances of the observed time series at A23–A29 are small.

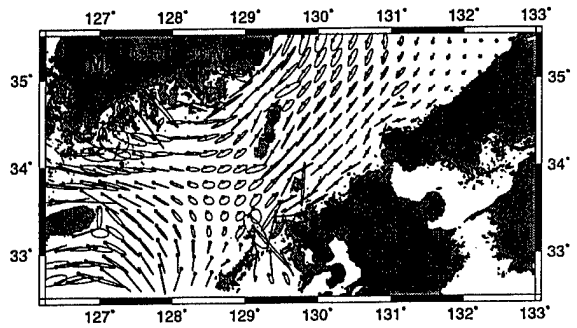

Fig. 7. Model predicted tidal current ellipses for the M_2 tide. Linear interpolation was used over the model's C grid in order to collocate u and v predictions. For display purposes, every fourth ellipse is drawn. Three ellipses in the northwest Jeju Strait and one ellipse in the narrows between the Gotou Islands and Kyushu are drawn at half scale with a dashed line. Dots are drawn to indicate the tip of the tidal current vector at the time when the astronomical argument is zero (Foreman, 1978). Stars are drawn in place of dots for the half scale ellipses.

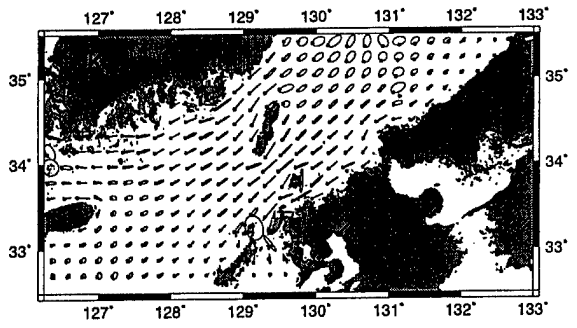

Fig. 8. As Fig. 7, but for the K_1 tide. All ellipses are drawn with the same scale.

4.2 *Model predictions*

Figures 3–12 present the model tidal predictions throughout the model domain. The fields were made by harmonically analyzing the model's time domain output. Nodal corrections have been applied to all amplitudes and phases. Fields for M_2, S_2, K_1, and O_1 were generated from the four-constituent model run; those for N_2, P_1, K_2, and μ_2 were generated from the eight-constituent model run. The Rayleigh criterion is violated for the K_1/P_1 and the S_2/K_2 pairs in the harmonic analysis of this latter run because of the 41-day duration of the model time series. However this does not increase the error for these constituents because the background, i.e. non-tidal, energy from the model is negligible. Tidal amphidromes are marked in these figures if the amphidrome is completely enclosed by the model domain.

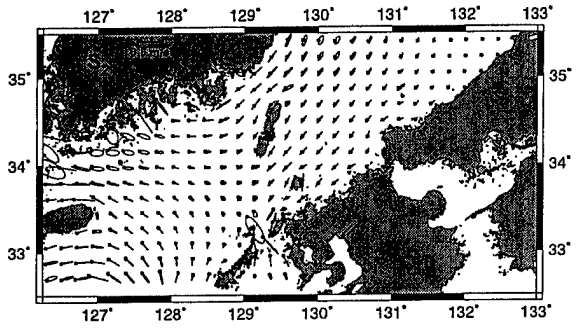

Fig. 9. As Fig. 7, but for the S_2 tide. All ellipses are drawn with the same scale.

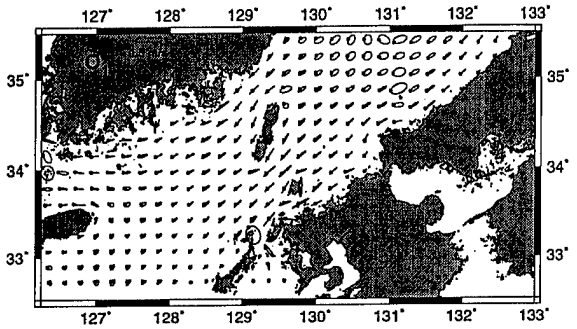

Fig. 10. As Fig. 7, but for the O_1 tide. All ellipses are drawn with the same scale.

4.3 *Model prediction accuracy*

The model provides tidal predictions at all grid points as well as at the observation locations. How well does the model represent the true barotropic tide throughout its domain? For the purposes of further error analysis it is convenient to express the data assimilation model in a matrix form, as was done by Dowd and Thompson (1996) and Thompson *et al.* (2000). Thompson *et al.* (2000) showed that for one tidal constituent, a form of the model can be written as

$$\mu_D = R_D \beta, \tag{10}$$

where μ_D is a column vector containing sine and cosine amplitude coefficients from the harmonically analyzed model output at all observation locations, R_D is a response matrix, and β is a column vector containing sine and cosine amplitude coefficients for all the structure functions from the harmonically analyzed forcing time series (F from Eq. (7)). Following the same methods presented in the appendices of Dowd and Thompson (1996), another form of the model can be written as

$$\mu_G = R_G \beta, \tag{11}$$

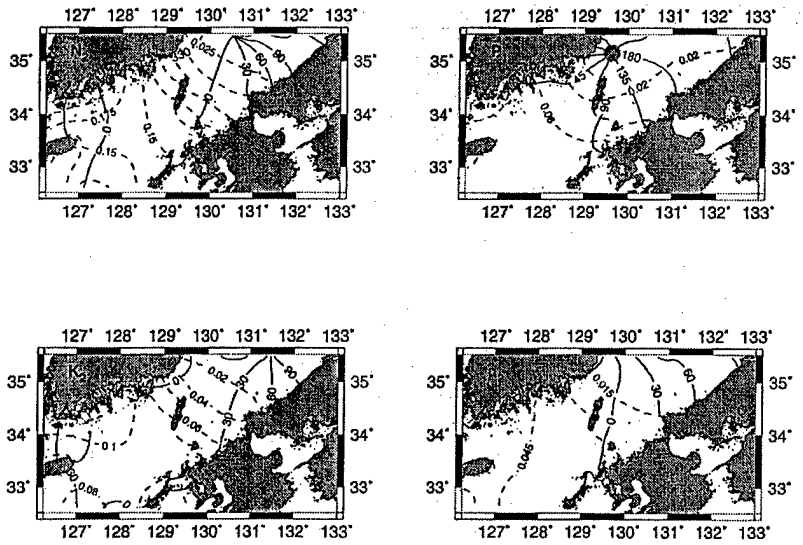

Fig. 11. As Fig. 3, but for the N_2, P_1, K_2, and μ_2 tides. The co-range lines are in increments of 0.025 m, 0.020 m, 0.020 m, and 0.015 m respectively. The co-tidal lines are in increments of 30 degrees for N_2, K_2, and μ_2 and in increments of 45 degrees for P_1. The gray circle indicates the position of the P_1 amphidromic point.

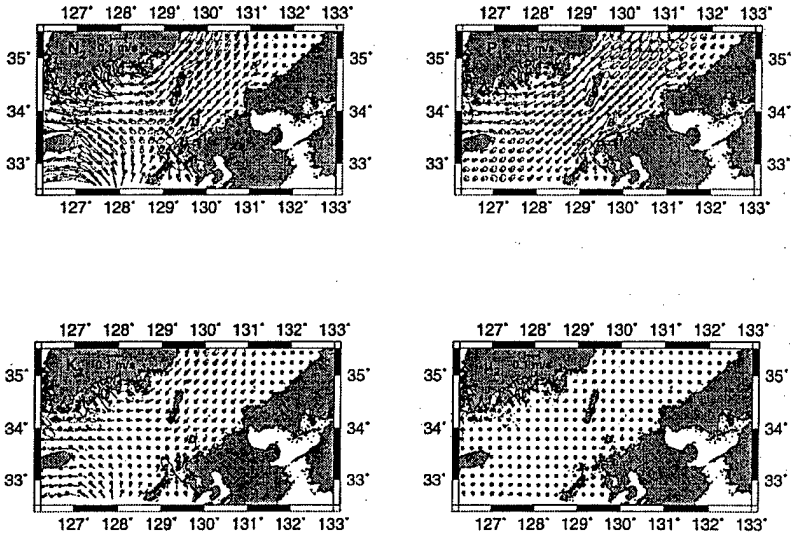

Fig. 12. As Fig. 7, but for the N_2, P_1, K_2, and μ_2 tides. The ellipses are all at 5 times the scale used in Figs. 8–10.

where μ_G is a column vector containing sine and cosine amplitude coefficients from the harmonically analyzed model output at all grid points, R_G is a different response matrix, and β is as defined above. Equations (10) and (11) represent the forward model for a particular tidal constituent. The actual model is run in the time domain as described in Section 3. Solutions for R_D and R_G were obtained from time domain runs of the forward model by following the alternative method described at the end of appendix A in Dowd and Thompson (1996). This method forces the model with a unit amplitude sine or cosine time series at a single structure function and zero at all other structure functions. In matrix form, one value of β has been set to one and all other values set to zero. The harmonically analyzed output from such a run yields a particular column of R_D and a particular column of R_G. Further columns of R_D and R_G were obtained from additional forward model runs by changing the structure function used for forcing.

Thompson *et al.* (2000) state the matrix form of their

unweighted cost function and then calculate the optimal choice for β and the expected variance of this optimal choice using standard linear regression techniques. Using our weighted cost function (Eq. (8)), the matrix equations for the cost function, the optimal choice for β, and the expected variance of this optimal choice are

$$J = (m - \mu_D)' w (m - \mu_D), \tag{12}$$

$$\beta_o = (R_D' w R_D)^{-1} R_D' w m, \tag{13}$$

and

$$\mathrm{var}(\beta_o) = (R_D' w R_D)^{-1} R_D' w \,\mathrm{var}(m) w R_D (R_D' w R_D)^{-1}, \tag{14}$$

where m is a column vector containing the sine and cosine amplitude coefficients for the harmonically analyzed observation time series, β_o is the optimal choice for β, and w is a diagonal matrix containing weighting values derived from the values of W in Eq. (8). Primes indicate the transpose operation. The notation var(·) will be used throughout the paper to indicate the ensemble variance of a quantity.

For the model used in this study, there are particular values of β such that μ_D in Eq. (10) is approximately zero. The amplitudes and the error variances of the amplitudes for such singular vectors will be ill determined in Eqs. (13) and (14) and reach high values. Nevertheless, final amplitudes of these ill-determined vectors in the time domain model remain low because of the initialization of the vectors to zero and the conjugate gradient methodology. Therefore, the matrix model analog requires a special treatment of these vectors to accurately reflect the time domain model and to prevent unrealistic forcings from affecting the error estimates.

To accomplish this, first the singular vectors of R_D were found by a singular value decomposition,

$$R_D = U \Lambda V'. \tag{15}$$

V is a unitary matrix with orthonormal columns that are the singular vectors of R_D. These vectors are ranked by row according to the amount of response produced at the data sites by using them as β in Eq. (10). The response levels are quantified by the diagonal of Λ and the structure of the responses are given by the columns of U. Converting the problem to the solution space spanned by the singular vectors gives

$$R_D^{\ddagger} = R_D V = U \Lambda. \tag{16}$$

Now the solution space can be divided into the determined solution space and the null space. For this study, the dividing criteria classified a singular vector in the null space if all responses to the vector at the data sites were less than 1% of the vector amplitude, i.e. the corresponding column of $R_D^{\ddagger}$ had values all less than 0.01. For M_2, 18 of the singular vectors were classified in the null space and 38 vectors were classified in the determined space. After this separation, the null space can be truncated from the solution by defining a matrix R_D^* that is made up of the first 38 columns of $R_D^{\ddagger}$ and a matrix $R_D^{\dagger}$ that is made up of the last 18 columns of $R_D^{\ddagger}$. In this new space, the variances of the determined singular vectors, β^*, are found by applying Eq. (14) using R_D^* instead of R_D. The variances of the null singular vectors, $\beta^{\dagger}$, are set to one and their co-variances are set to zero. Large values for the amplitudes of these vectors would either produce detectable changes at the data sites or lead to unrealistically large tidal fluctuations elsewhere in the model domain and therefore cannot occur in nature.

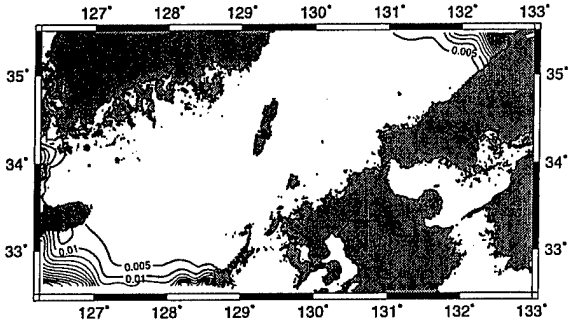

Fig. 13. Estimated error in the predictions of M_2 η tidal amplitudes from the deterministic part of the solution space (see Subsection 4.3). Contours are drawn at 0.005 m increments. The 0.005 and 0.01 m contours are drawn as bold lines

In Eq. (11), the tidal predictions at the grid points are determined by a constant term multiplied by β_o. Therefore, in the new space, the expected variance matrices for these predictions are

$$\mathrm{var}(\mu_G^*) = R_G V^* \left(R_D^{*\prime} w R_D^* \right)^{-1} R_D^{*\prime} w \,\mathrm{var}(m) w R_D^* \left(R_D^{*\prime} w R_D^* \right)^{-1} V^{*\prime} R_G' \tag{17}$$

and

$$\mathrm{var}(\mu_G^{\dagger}) = R_G V^{\dagger} \,\mathrm{var}(\beta^{\dagger}) V^{\dagger\prime} R_G', \tag{18}$$

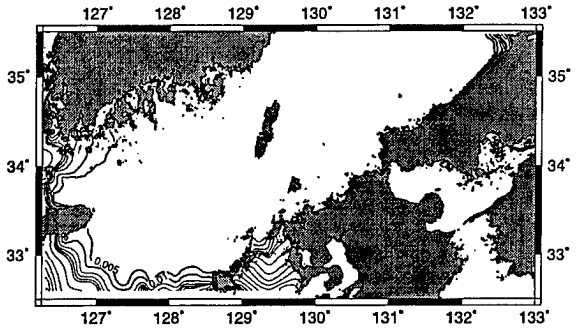

Fig. 14. Estimated error in the predictions of M_2 η tidal amplitudes caused by the undetermined singular vectors (see Subsection 4.3). The 0.005 and 0.01 m contours are drawn as bold lines. The 0.02, 0.03, 0.04, and 0.05 m contours are drawn with normal thickness. Additional contours are drawn at 0.1 m increments.

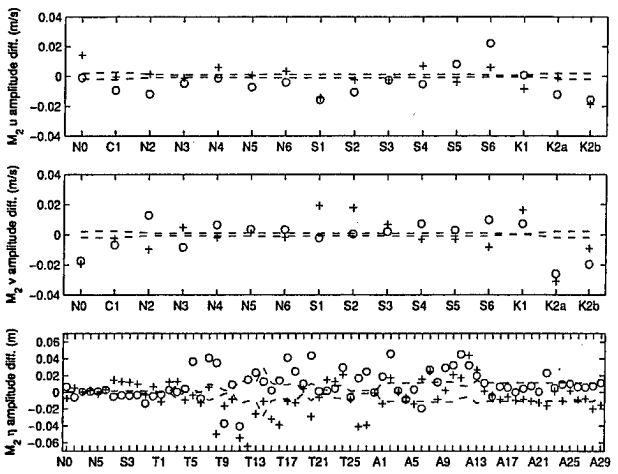

Fig. 15. Differences between the observed and modeled M_2 cosine (circles) and sine (pluses) amplitudes at all the data sites. Dashed lines are drawn ± one standard deviation from the expected value (zero) of these differences. The positions of these lines were determined from Eq. (22).

where V^* and $V^\dagger$ are the first 38 columns of V and the last 18 columns of V respectively. Figures 13 and 14 present the predicted M_2 η errors for the four-constituent run resulting from the use of Eqs. (17) and (18). The predicted errors for most of the model domain are less than 0.5 cm. The model null space (Fig. 14) occurs along the far southwestern coast of Korea, south of Jeju Island, between the Gotou Islands and Kyushu, and along the coast of Honshu north of 35.2°. The errors shown in this figure are true only if the amplitudes of the null space vectors can in actuality reach values of one.

4.4 Model accuracy

The theory in Subsection 4.3 has assumed that the model physics, numerics, and parameters are true. Under these assumptions the error in the model over most of the domain is very small. How large are the errors in the model due to imperfect model physics, numerics, and parameters? Figure 2 suggests that these errors may be larger than the errors calculated from Eq. (17).

For ease of notation, let

$$S = R_D^*\left(R_D^{*\prime} w R_D^*\right)^{-1} R_D^{*\prime} w. \tag{19}$$

Using this expression and combining Eqs. (10) and (13), the measured M_2 amplitude values minus their model counterparts are then

$$m - \mu_D = m - Sm = (I - S)m. \tag{20}$$

In this equation, I is the identity matrix. Using the truncated form of R_D in Eq. (19) allows the accurate evaluation of the matrix inverse and does not drastically affect the results because the null space vectors have small projections to the data sites. If m is composed of random noise plus a signal that satisfies the model physics, numerics, and parameters, then the expected value, E(·), and variance of Eq. (20) are

$$\mathrm{E}(m - \mu_D) = 0 \tag{21}$$

and

$$\mathrm{var}(m - \mu_D) = \mathrm{var}(m) - S\,\mathrm{var}(m) - \mathrm{var}(m)S' + S\,\mathrm{var}(m)S'. \tag{22}$$

In Eq. (22), the second and third terms are equal because S is symmetric and var(m) is diagonal. The magnitude of the fourth term is also equal to the magnitude of these terms if the weighting in the model is optimal. var(m) represents the random error in the data at each site and $-S$var(m) represents over fitting of the model. The combined term of $-$var(m)S' + Svar(m)S' represents the affect of weighting some sites too strongly and others too weakly.

Figure 15 shows the difference between the measured M_2 sine and cosine amplitudes and the predicted M_2 amplitudes for the four-constituent run. Values ± one standard deviation from the expected value (zero) of these differences, calculated from Eq. (22), are plotted as dashed lines. If the theory presented by Eqs. (21) and (22) is correct then the chance is small that the differences in Fig. 15 would be far above or below one standard deviation.

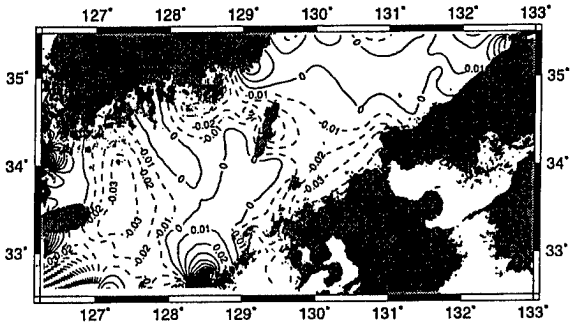

Fig. 16. Estimated bias in the predictions of M_2 η tidal amplitudes. Contours are drawn at 0.01 m increments. Negative values (dashed contours) indicate that the predicted model amplitudes are too low. These error values were derived from Eq. (25).

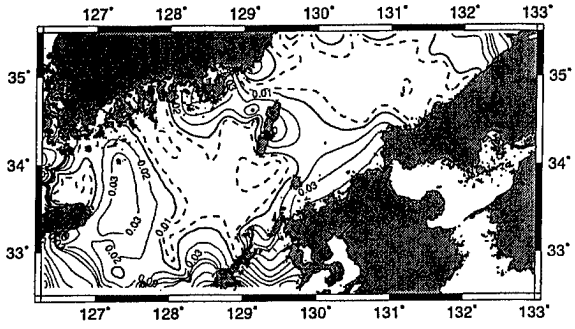

Fig. 17. The total estimated rms error for the M_2 η tidal amplitudes. These errors are equal to the square root of the sum of the squares of the errors in Figs. 13, 14, and 16. The 0.005 m contour is drawn as a dashed line and the 0.01 m contour is drawn as a bold line. The 0.02, 0.03, 0.04, and 0.05 m contours are drawn with normal thickness. Additional contours are drawn at 0.1 m increments.

Statistically, this theory can be rejected at the 99% confidence level at all the velocity sites and at 54% of the η sites. Most of the sites where the theory cannot be rejected based on the evidence are T/P measurement locations.

There are two possible explanations for the failure of the theory. The first is that the error variances estimates, var(m), are inaccurate and the standard deviation lines in Fig. 15 should be farther away from zero. The second is that m is composed of more than random noise plus tidal signals that follow the linear model physics and model parameters, i.e. the model physics, numerics, and/or parameters contain error. The second explanation is more likely because of the large degree of departure from the error theory at many sites where the error variance estimates should be very accurate, and because of the many possible sources of error including bathymetry errors, error in the friction formulation, error in the friction coefficient, resolution of the structure functions, non-linear tide interactions, etc. Therefore, this study will assume that the error variance estimates are accurate and that the larger errors are due to errors in the model, or errors in the model parameters.

To account for this, let m now be composed of a signal that satisfies the model, plus a true tide signal, N_D, that is not represented in the model, plus random noise. With this new theory, the variance of $(m - \mu_D)$ is still given by Eq. (22) but Eq. (21) becomes

$$\mathrm{E}(m - \mu_D) = (I - S)N_D. \tag{23}$$

Although the data provide only a single realization of $(m - \mu_D)$, this single realization can be used as an accurate estimate of the expected value because the variance of this quantity is low (Fig. 15). The matrix $(I - S)$ is always singular because of the structure of S and therefore Eq. (23) cannot be simply inverted to solve for the unknown values of N_D. However, an approximation to N_D was found by solving the related equation

$$(I - S)'(m - \mu_D) = (I - S)'(I - S)N_D \tag{24}$$

using the conjugate gradient method (Golub and Van Loan, 1996).

The bias of μ_G is

$$\mathrm{bias}(\mu_G) = R_G V^* \left(R_D^{*\prime} w R_D^* \right)^{-1} R_D^{*\prime} w N_D - N_G, \tag{25}$$

where N_G is the true tide signal at the model grid points that is not represented by the model. N_G was estimated by gridding (Smith and Wessel, 1990) the values of N_D and then Eq. (25) was used to estimate the bias in μ_G. Figure 16 shows the M_2 amplitude bias estimate derived from the results of Eq. (25). Negative values (dashed contours) indicate that the predicted model amplitudes are too low. In general the M_2 bias field is between –2 and 1 cm, but it reaches values less than –3 cm to the west of Busan, around Tsushima Island, to the east of Jeju Island, and along the coast of Kyushu. Values higher than 2 cm occur directly to the west of the Gotou Islands and in the null space south and north of Jeju Island.

Figure 17 shows the estimated rms error for the M_2 amplitude field. This map combines the errors from Figs. 13, 14, and 16, but the errors from Fig. 16 dominate in most locations. For broad regions near the north and south LINKS mooring lines and in the Jeju Strait the rms errors are less than 0.5 cm. Elsewhere, outside the null space, errors range from 1 to 4 cm.

5. Discussion

Figures 3–12 show the complex structures and details of the barotropic tide in the Korea/Tsushima Strait. The tidal response in the strait is strongly frequency dependent. In particular, the semi-diurnal and diurnal constituents exhibit distinct group patterns in their surface heights and currents. Fang and Yang (1988) identified the energy inflowing regions of the strait to be west of the Gotou Islands and in the north near Korea and the outflowing regions to be in the Jeju Strait and in the north near Honshu. In these various regions the relative importance of the semi-diurnal and diurnal tides change with respect to each other, thus representing frequency-dependent changes induced as the tides transverse the strait. Significant differences also occur in the basin response to O_1 compared to the response to K_1.

The M_2 tide has high amplitudes and dominates tidal surface height fluctuations throughout the strait. A broad area southwest of Tsushima Island exists where the semidiurnal tide changes little in amplitude or phase. The surface heights of this region all oscillate together. The model reproduces the 10 cm semidiurnal height amplitude difference from the west coast to the east coast of Tsushima Island that was observed by Odamaki (1989). Amphidromic points for the M_2 and S_2 tides appear to lie just outside of the northern model boundary.

Although the model semidiurnal tide charts are similar in character to those of Odamaki (1989), significant differences exist. Most notable is the difference in amplitude along the south coast of Korea, where our predictions are lower by approximately 10 cm than those of Odamaki (1989). Possibly, the historical charts are biased high because of their reliance on coastal data where tidal fluctuations are often enhanced. Differences also occur south of Jeju Island, but this region is near the model null space (Fig. 14) and the model may not be able to predict the total tide signal there.

The diurnal tidal heights are, in general, smaller than those of the M_2 and S_2 tides. However, in the northeast portion of the strait, the semi-diurnal amplitudes rapidly decrease toward the northeast while the diurnal amplitudes remain relatively constant and thus the amplitudes of the O_1 and K_1 tides exceed those of S_2. The K_1 height fluctuations are stronger than those of O_1 for the half of the strait west of Tsushima Island and slightly weaker for the half east of Tsushima Island. The model predicts the 2.5–5 cm diurnal height difference from the east to the west coast of Tsushima Island observed by Odamaki (1989). Our predictions place the diurnal amphidromic points inside the model domain. Their positions do not occur near the null space and therefore should be accurate. The location of the amphidromic point for K_1 is different between our model and the charts of Odamaki (1989). His charts place the K_1 amphidrome against the Korean coast just south of the city of Ulsan (tidal station T10), whereas our model shows the K_1 amphidrome located offshore and farther southward.

M_2 currents are strong throughout the strait and dominate the tidal currents in most areas. Southwest of Tsushima Island, the M_2 ellipses are weaker and more circular than elsewhere. In this region, the major semiaxes of the K_1 and the O_1 currents are approximately the same as those of M_2. In general, the strength of the S_2, K_1, and O_1 currents are similar. S_2 currents are stronger than either K_1 and O_1 currents north and south of Jeju Island, along the southern boundary, and near Busan but weaker almost everywhere else. For a broad region in the northeast portion of the strait, O_1 and K_1 currents exceed all other constituents. Here, the O_1 currents are also significantly larger than the K_1 currents and thus O_1 is the dominate tidal current constituent for this region offshore of Honshu.

Figures 11 and 12 show the tides for the next four largest components. Although the height fluctuations of N_2 are of similar magnitude to that of K_1 and are often greater than that of O_1, the tidal currents for this component are much smaller than the four main components. Likewise K_2 has greater tidal height fluctuations than P_1 but smaller tidal currents. The μ_2 constituent is of little significance in the Korea/Tsushima Strait.

Kang (1984) and Lee and Kim (1999) have shown that the tides near Jeju Island exhibit complex local effects that are difficult to model. In contrast to the increasing tidal height fluctuations from south to north near Jeju Island, the tidal height fluctuations are stronger on the south side of the island than on the north side of the island. Thus the observed M_2 and K_1 amplitudes are 70 and 23 cm at station T2, but are 77 and 24 cm at station T1. Our model successfully predicts this increase, giving amplitudes of 69 and 23 cm at station T2, and 77 and 24 cm at station T1. Correctly weighting the cost function according to the expected error in the observational dataset was required to obtain this result. With correct weighting the model is able to find a solution that matches the observations within the restrictions of strictly linear dynamics.

The model prediction errors (Fig. 17) are small relative to the amplitudes of the tidal signals in the Korea/Tsushima Strait (Fig. 3) and are dominated by bias error in the model's physics, numerics, and/or parameterization. The model can only use linear dynamics to propagate incoming waves from the boundary to match observations. The differences shown in Fig. 15 indicate the possible strength of non-linear effects in determining the true barotropic tide in the strait. However, it is likely that some portion of these differences is not caused by non-linear effects but rather by model parameter errors such as errors in the bathymetry field. In particular, bathymetric

sensitivity runs made in this study show that localized changes in the model bathymetry can degrade the model fit over broad spatial areas. Odamaki (1989) showed that the tide-generating force in the strait is very small compared to the other terms in the dynamical equations. Our model agrees with this finding. If direct forcing of tides were important, differences between the observations and the model predictions would have been greater.

Model prediction errors peak sharply near portions of the model boundary (Fig. 14). The model currents at these boundaries are determined by the observations through the inverse method described in this paper. However, the error results do not suggest that the forcings so determined are invalid. The model determines with a high degree of accuracy the portion of the boundary flows that impact the data sites. Additional flows could exist in these regions that lead to no impact at the data sites. The model cannot determine if such flows exist. Therefore, the high errors near the model boundary represent the possibility that the true barotropic tide could consist of flows that do not affect the data sites.

6. Conclusions

The entire Korea/Tsushima Strait has high tidal currents, with several constituents having significant amplitudes. By using an extensive set of measurements inside the strait to predict boundary flows for a numerical model in a strong-constraint inverse problem, tidal predictions throughout the strait were obtained. The predictions follow strictly linear dynamics and provide a best match in a least-squares sense to the observations, weighted according to their accuracies. Model bias was estimated by comparing model-observation differences and the theoretical expected values and ensemble variances of these differences calculated only from the model dynamics and measurement accuracies. The total estimated rms errors, including model bias, for M_2 η predictions are less than 3 cm over most of the model domain.

The tides in the Strait exhibit a complex structure in which diurnal constituents have stronger currents relative to their sea level height ranges than semi-diurnal constituents. No single constituent dominates both tidal sea level heights and velocities in all parts of the Strait. Tidal constituents M_2, S_2, K_1, O_1, and N_2 have the highest ranges in sea level height. Constituents M_2, K_1, S_2, and O_1 have the highest magnitudes of tidal velocities.

Acknowledgements

This work was sponsored by the Office of Naval Research as part of the basic research project "Linkages of Asian Marginal Seas" under program element 0601153N (NRL-SSC contribution JA/7330/03/17). K. I. Chang and M. S. Suk were supported by grants from KORDI's in-house project "Marine Ecosystem Response to Climate variability in the East Sea and Tecnotic Evolution." J. C. Lee was supported by the Korea Research Foundation in 1999. We would like to thank R. Burge, M. Hulbert, and D. Kennedy for their work in the deployment and recovery of the LINKS moorings. Dr. Youyu Lu provided help with the data assimilation model during the summer of 2000 at Dalhousie University, Dr. Mark Wimbush provided helpful comments about the manuscript text, and Dr. Hans Ngodock greatly helped with the matrix algebra in the error analysis. The comments of two anonymous reviewers led to significant improvement of this study.

References

Choi, B. H. (1999): *Digital Atlas for Neighboring Seas of Korean Peninsula.* Laboratory for Coastal and Ocean Dynamics Studies, Sung Kyun Kwan University, CD-ROM.

Choi, B. H., D. H. Kim and Y. Fang (1989): Tides in the East Asian Seas from a fine-resolution global ocean tidal model. *MTS Journal*, **33**, 36–44.

Dowd, M. and K. R. Thompson (1996): Extraction of tidal streams from a ship-borne acoustic Doppler current profiler using a statistical-dynamical model. *J. Geophys. Res.*, **101**(C4), 8943–8956.

Fang, G. and J. Yang (1988): Modeling and prediction of tidal currents in the Korea Strait. *Prog. Oceanogr.*, **21**, 307–318.

Foreman, M. G. G. (1977): Manual for tidal heights analysis and prediction. Pacific Marine Science Report, 77-10, Institute of Ocean Science, Patricia Bay, Sidney, B.C., Canada.

Foreman, M. G. G. (1978): Manual for tidal currents analysis and prediction. Pacific Marine Science Report, 78-6, Institute of Ocean Science, Patricia Bay, Sidney, B.C., Canada.

Golub, G. H. and C. F. Van Loan (1996): *Matrix Computations: third edition.* Johns Hopkins University Press.

Griffin, D. A. and K. R. Thompson (1996): The adjoint method of data assimilation used operationally for shelf circulation. *J. Geophys. Res.*, **101**(C2), 3457–3477.

Hallock, Z. R., P. Pistek, J. W. Book, J. L. Miller, L. K. Shay and H. T. Perkins (2003): A description of tides near the Chesapeake Bay entrance using in situ data with an adjoint model. *J. Geophys. Res.*, **108**(C3), doi:10.1029/2001JC000820.

Kang, S. K., S.-R. Lee and K.-D. Yum (1991): Tidal computation of the East China Sea, the Yellow Sea, and the East Sea. p. 25–48. In *Oceanography of Asian Marginal Seas*, ed. by K. Takano, Elsevier.

Kang, S. K., J.-Y. Chung, S.-R. Lee and K.-D. Yum (1995): Seasonal variability of the M2 tide in the seas adjacent to Korea. *Cont. Shelf Res.*, **15**(9), 1087–1113.

Kang, Y. Q. (1984): An analytic model of the M2 tide near Cheju Island. *The Journal of the Oceanological Society of Korea*, **19**(1), 18–23.

Kantha, L. H., I. Bang, J.-K. Choi and M.-S. Suk (1996): Shallow water tides in the seas around Korea. *The Journal of the Korean Society of Oceanography*, **31**(3), 123–133.

Lee, S. and K. Kim (1999): Scattering of tidal frequency waves around a cylindrical island. *J. Phys. Oceanogr.*, **29**, 436–448.

Lefèvre, F., C. Le Provost and F. H. Lyard (2000): How can we improve a global ocean tide model at a regional scale? A test on the Yellow Sea and the East China Sea. *J. Geophys. Res.*, **105**(C4), 8707–8725.

Naval Oceanographic Office (1997): Data Base Description for Digital Bathymetric Data Base—Variable Resolution (DBDB-V) Version 1.0, Internal Report, Naval Oceanographic Office, Stennis Space Center, MS, U.S.A.

Odamaki, M. (1989): Tides and tidal currents in the Tusima Strait. *J. Oceanogr. Soc. Japan*, **45**, 65–82.

Ogura, S. (1933): The tides in the Seas adjacent to Japan. *Bulletin of Hydrographic Department*, **7**, Imperial Japanese Navy.

Reid, R. O. and B. R. Bodine (1968): Numerical model for storm surges in Galveston Bay. *Journal of the Waterways and Harbors Division, Proceedings of the American Society of Civil Engineers*, **94**(WW1), 33–57.

Smith, W. H. F. and P. Wessel (1990): Gridding with continuous curvature splines in tension. *Geophysics*, **55**(3), 293–305.

Teague, W. J., P. Pistek, G. A. Jacobs and H. T. Perkins (2000): Evaluation of tides from TOPEX in the Bohai and Yellow Sea. *J. Atmos. Ocean. Technol.*, **17**, 679–687.

Teague, W. J., H. T. Perkins, G. A. Jacobs and J. W. Book (2001): Tide observations in the Korea-Tsushima Strait. *Cont. Shelf Res.*, **21**, 545–561.

Teague, W. J., H. T. Perkins, G. A. Jacobs, J. W. Book, K.-I. Chang and M.-S. Suk (2002): Low frequency current observations in the Korea/Tsushima Strait. *J. Phys. Oceanogr.*, **32**, 1621–1641.

Thacker, W. C. (1987): Three lectures on fitting numerical models to observations. Report GKSS 87/E/65, GKSS Forschungszentrum Geesthacht GmbH, Geesthacht, Germany.

Thompson, K. R., M. Dowd, Y. Lu and B. Smith (2000): Oceanographic data assimilation and regression analysis. *Environmetrics*, **11**, 183–196.

Wessel, P. and W. H. F. Smith (1998): New, improved version of Generic Mapping Tools released. *EOS, Trans., A.G.U.*, **79**(47), 579.