

NCEP NOTES

Virtual Floe Ice Drift Forecast Model Intercomparison*

ROBERT W. GRUMBINE

National Centers for Environmental Prediction, Camp Springs, Maryland

5 September 1996 and 10 March 1998

ABSTRACT

Both sea ice forecast models and methods to measure their skill are needed for operational sea ice forecasting. Two simple sea ice models are described and tested here. Four different measures of skill are also tested. The forecasts from the newer sea ice model are found to perform better, regardless of the skill measure used. All four skill measures show essentially the same behavior, in terms of having no dependence on season and being roughly constant. All four measures also agree that there is no decline in skill with time through the 6-day period of forecast.

1. Introduction

Since at least the time of Nansen (1902), it has been common to think of sea ice drifting at some fraction of the wind speed, and at some angle to the wind. This is the drift rule. Nansen's values, based on observation of ice floe drift during the cross-polar drift of the *Maud* (1893–96) were 1.8% and 28° to the right of the wind. This included about 949 floe-days of observations (7 November 1893–27 June 1896) from a single point.

In the time since then, two things have changed for drift models. It has become the convention to use geostrophic winds rather than surface winds in deriving the drift law. And the number of observations has increased dramatically. Thorndike and Colony (1982) analyzed 7937 buoy-days of observations. Their simple drift law (0.8%, 8° to the right of the geostrophic wind) was able to explain 70% of the variance in drift velocity in the central Arctic basin. In the Antarctic, Martinson and Wamser (1990) derived a drift law of 3%, 23.4° to the left of the geostrophic wind, from three points observed for 4–5 days each.

The United States has been running operationally a drift law forecast model since about 1968 (Skiles 1968), first at the Naval Oceanographic Office, then at the National Centers for Environmental Prediction formerly the National Meteorological Center). This

drift law was based on about 1080 buoy-days of observations. The forecast is used by the Anchorage Weather Service Forecast Office and the National Ice Center (NIC) in making their forecasts of ice edge position. Their forecasts are then used by shipping companies, fishermen, and oil companies. We present evidence that suggests that the forecast has been improved by using more recent drift laws, and by adding the Southern Hemisphere (not forecast in the formerly operational model, but forecast by the NIC).

The sea ice literature includes relatively little in the way of quantitative model verification. For the most part, this has been because visual inspection of model output has been sufficiently unambiguous to determine which model or parameterization was better. Skill measurements that have been used have included ice drift distance correlation between forecast and observation (Ip 1991; Flato and Hibler 1992) and index of agreement between forecast and observed ice drift (Preller and Posey 1989). Neither of these is a vector measure, so that forecasting the right distance in the wrong direction is still credited as a good forecast. In addition to these two, the vector correlation definition from Crosby et al. (1993) and error radius (Flato and Hibler 1992) are tested. These four measures of skill will be examined by themselves, as well as used to determine which model is better.

Also to be examined is the dependence of forecast skill on length of forecast. Contrary to the more common differential models (cf. weather, waves, sea surface temperature) where *instantaneous* values of variables governed by (partial) differential equations are desired, ice drift is an integral model. The forecast quantity is the total drift (distance and direction) *integrated* over the whole forecast period. Over the first 6 days of the fore-

* OMB Contribution Number 114.

Corresponding author address: Dr. Robert W. Grumbine, NCEP, 5200 Auth Road, Camp Springs, MD 20746.
E-mail: seaic@polar.wwb.noaa.gov

cast period (the maximum length run), there is no systematic decline of skill (by any of the four measures) with respect to time. This counterintuitive point will be discussed in some length in section 3.

2. Ice drift models

The models discussed are based on a virtual floe concept. The prediction is how far and in what direction a floe would drift, if there were a floe at a given point to start with, and if it is assumed that it does not melt or encounter coasts. Forecasters must then temper the model output with their knowledge of meteorological, oceanographic, and coastal effects. For all the drift rules, the two constants to be determined are the ice drift speed and the drift direction relative to the geostrophic wind speed and direction.

The formerly operational model was developed by Skiles (1968) for the Naval Oceanographic Office. The drift rule, based on an examination of buoy data (Skiles 1968), is

$$U_i = (5.36 \times 10^{-4} + 7.72 \times 10^{-3}|U_a|) \quad (1)$$

$$\theta_i = \theta_a + 31.3 \exp(-3.60 \times 10^{-3}|U_a|), \quad (2)$$

where U_i is the ice velocity, U_a is the geostrophic wind velocity, velocities are in meters per second, θ_a is the wind direction, and θ_i represents the iced drift direction. The operational implementation of the Skiles model also included a process to “roughen” the pressure field prior to determining the geostrophic wind. No reason for doing this was documented in the model code, and the process is not part of the description given by Skiles (1968). The procedure is

$$P'^j_i = D_0 P^j_i - D_1 (P^j_{i+1} + P^j_{i-1} + P^{j+1}_i + P^{j-1}_i) + D_2 (P^{j+1}_{i+1} + P^{j-1}_{i-1} + P^{j+1}_{i+1} + P^{j-1}_{i-1}), \quad (3)$$

where i, j are model grid points, P'^j_i are the roughened pressures, P^j_i are the model-derived pressures, and D_0, D_1, D_2 are assigned 16/9, 2/9, and 1/36, respectively. This is not, strictly speaking, a Laplacean operation as the metric terms of the polar stereographic grid are ignored in this step. The operational Skiles model used the 2.5° gridded forecast pressure fields.

The newer model is based on the drift rule of Thorndike and Colony (1982) for the Northern Hemisphere, and Martinson and Wamser (1990) for the Southern. In the Northern Hemisphere,

$$U_i = 0.008 \overline{\mathbf{R}}(\theta) U_a, \quad (4)$$

where $\overline{\mathbf{R}}$ is the two-dimensional rotation matrix:

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad (5)$$

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \quad (6)$$

and θ is 8°.

In the Southern Hemisphere, the rule is

$$U_i = 0.03 \overline{\mathbf{R}}(\theta) U_a, \quad (7)$$

where θ is -23° . The ice drifts more rapidly and at a greater angle because Antarctic ice is generally thinner than Arctic. The newer model computes the geostrophic winds in the spectral domain then interpolates them to a 1° mesh. No roughening procedure is used.

3. Model intercomparison

The results of the two models' forecasts were compared by objective verification against observed buoy drifts in the Arctic, and subjectively by the Anchorage forecast office and the NIC. The Anchorage office (C. Bauer 1994, personal communication; R. Page 1995, personal communication) states that the revised model is indeed superior to the Skiles model. The differences were said to be particularly notable for low drift conditions (an observation the objective verification seconds). The NIC (D. Helms 1995, personal communication) finds the Thorndike and Colony implementation to be superior in the Arctic generally. For the Antarctic this implementation is the only one available and is considered helpful (D. Helms 1995, personal communication).

We take correct forecasting of the drift distance (and potentially, direction) integrated through the length of the forecast as the measure of success. This is the term that operational forecasters are concerned with. We also consider the skill as a function of time in the forecast model. In scoring the models, we also want to look for scoring measures that differentiate most strongly between the models. Previous measures suffer from the problem that even large model differences can result in small forecast score differences (Ip et al. 1991; Flato and Hibler 1992; Preller and Posey 1989).

The forecast ice drifts are verified against the observed buoy drift for each forecast day. The comparison point was the virtual floe point closest to the starting location of the buoy each day during the forecast period. The floe point and buoy were required to be within 55 km initially for comparisons to be made. No interpolation was done. The initial time (at which the forecast is started and the time that the buoy position is checked) is 0000 UTC. The position of a virtual floe starting from a given location is made for final times every 12 h to 6 days. Forecast values are the location after N hours of a floe that had started at the given location. If there were multiple buoy reports within 3 h of 0000 UTC, the average location was assigned to 0000 UTC.

The four measures of skill are the correlation of distance, the index of agreement in distance (Willmott et al. 1985), error radius (position location error between forecast and observed location), and vector correlation in drift after Crosby et al. (1993). Correlation varies from -1 to 1 , index of agreement from 0 to 1 , and vector correlation from 0 to 2 .

Scalar correlation, the correlation between a fore-

TABLE 1. Index of agreement (IA) for forecast drift distance, correlation of forecast and observed drift distance R, vector correlation of observed and forecast drift (VCC), and *t* statistic for error radius as a function of forecast length (day). Here, N is the number of buoy-days of observations available for verification. The formerly operational Skiles model is listed first in each pair. The *t* statistic is given in the sense that the new model is better (smaller errors) when negative, and asterisked when it is significant at the 95% level.

Day	N	IA	R	VCC	<i>t</i>
1	2204	0.464 0.515	0.455 0.496	0.602 0.595	-0.161
2	2069	0.475 0.528	0.489 0.526	0.653 0.674	-0.783
3	2024	0.474 0.526	0.484 0.535	0.650 0.685	-1.174
4	1875	0.463 0.516	0.464 0.520	0.636 0.688	-1.651*
5	1808	0.474 0.530	0.463 0.533	0.643 0.702	-1.561
6	1709	0.472 0.528	0.455 0.528	0.613 0.685	-2.600*

cast and observed parameter (drift distance in our case), is a standard measure of skill. It also has well-known failings (cf. Brier and Allen 1950). For our use, this includes that a consistent bias will be granted a high correlation. Also, an error of a given magnitude will be penalized the same, whether it is a 2-km error relative to an observation of 2 km, or an observation of 20 km.

The index of agreement may be interpreted as a relative average error (Willmott et al. 1985). It contrasts, then, with scalar correlation in that index of agreement will penalize a 2-km forecast error more if the observation is one of 2 km than if the observation is for 20 km. A bias will also be penalized. Mathematically, the formulation is (Willmott et al. 1985)

$$d_2 = 1 - (\sum \omega_j |\mathbf{e}_j|^2) / [\sum \omega_j (|\mathbf{p}_j - \bar{\mathbf{o}}| + |\mathbf{o}_j - \bar{\mathbf{o}}|)^2], \quad (8)$$

where d_2 is the index of agreement, summations are from 1 to N (the number of observations), \mathbf{e}_j is the error in the j th forecast, ω are weights to correct the \mathbf{e}_j for being over- or underrepresentative, \mathbf{p} is the prediction, \mathbf{o} is the observation, and $\bar{\mathbf{o}}$ is the mean of the observations weighted by ω . In our case, the weights are taken to be unity. The index of agreement will be largest when the numerator is smallest (forecasts agree with observations), and when the denominator is largest (large natural variability—the \mathbf{o}_j vary greatly from $\bar{\mathbf{o}}$).

The vector correlation defined by Crosby et al. (1993) is a generalization of the scalar correlation. As such, it involves terms like the covariance between predictions and observations scaled by the standard deviation of the predictions and the standard deviation of the observations. It has the usual properties of scalar correlation, including that if the predictions are a linear function of the observations, the correlation will be perfect (1 in a one-dimensional case, 2 in a two-dimensional case). As for scalar correlation, a 2-km forecast error will be considered equally bad regardless of whether the observed drift is 2 km or 20 km. The definition for a sample is

$$r^2 = \text{Tr}[\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}], \quad (9)$$

where Tr denotes the trace of the matrix and Σ_{ij} is

$$c(u_i, u_j) \quad c(u_i, v_j) \quad (10)$$

$$c(v_i, u_j) \quad c(v_i, v_j); \quad (11)$$

$c()$ is the sample covariance; $i, j = 1$ represent the observations, $i, j = 2$ are the forecasts; and u is drift distance in longitude, v is drift distance in latitude.

Model forecasts from 14 April 1993 to 31 January 1995 are scored. July 1994 is missing due to an archive failure. Various other days are also missing due, typically, to a computer queueing failure. The models' skill as a function of forecast length and skill measure is shown in Table 1 for all forecasts. For the error radius, the *t* statistic for the improvement (negative sign means improvement, i.e., drift location errors are smaller) is given. When significant at 95% level, it is starred. From Table 1, we clearly see that there is no notable relation between forecast length and skill, for any of the measures. The error radius shows an increase in statistical significance with time, exceeding 95% level for days 4 and 6 forecasts, and 90% at day 5. All other measures are essentially constants near the middle of their range. The new model appears to be consistently better than the old at all forecast intervals and for all scores (except day 1 in the vector correlation, where it is very slightly worse).

The absence of skill degradation with time is puzzling at first glance, since the atmospheric model that is used does become less skilled with time. The forecast variable is *net* drift over time. To be correct in forecasting this, it is only necessary to be correct in inferring the average velocity over the period of integration. The atmospheric model error can be considered composed of a bias and a random component. An atmospheric bias contributes to drift forecast error uniformly through time. The random component, however, can be expected to average out. It may be too high one day, but too low the next. This component contributes an error that will tend to decline with time. The process is that by which averaging many observations can lead to a more precise estimate of the mean that was obtainable from any single observation. The lack of a trend in skill with respect to forecast lead suggests that the loss of skill due to biases in the atmospheric model is approximately bal-

TABLE 2. Index of agreement (IA) for forecast drift distance, correlation of forecast and observed drift distance (R), vector correlation of observed and forecast drift (VCC), and *t* statistic of the difference in error radius between the two models for each month from April 1993 to January 1995 and asterisked when it is significant at the 95% level. Here, N is the number of buoy-days of observations available for verification. The operational Skiles model is listed first in each pair. The newer model is listed second.

Month/Year	N	IA	R	VCC	<i>t</i>
Apr 1993	16	0.580 0.640	0.522 0.652	0.598 0.683	-1.819*
May 1993	23	0.498 0.506	-0.068 0.013	0.456 0.439	-1.992*
Jun 1993	13	0.552 0.523	0.313 0.082	0.502 0.557	0.129
Jul 1993	66	0.689 0.597	0.451 0.345	0.812 0.779	-1.281
Aug 1993	96	0.454 0.513	0.148 0.288	0.683 0.732	-2.529*
Sep 1993	57	0.473 0.501	0.474 0.476	0.747 0.738	0.639
Oct 1993	129	0.447 0.475	0.316 0.372	0.603 0.652	0.410
Nov 1993	216	0.533 0.583	0.464 0.556	0.538 0.564	0.437
Dec 1993	120	0.471 0.479	0.254 0.236	0.589 0.464	2.110*
Jan 1994	64	0.523 0.645	0.644 0.749	0.636 0.743	-3.643*
Feb 1994	19	0.533 0.595	0.833 0.874	1.026 1.573	-4.510*
Mar 1994	88	0.379 0.430	-0.004 0.218	0.378 0.485	-5.162*
Apr 1994	75	0.563 0.590	0.616 0.581	0.958 0.910	0.491
May 1994	71	0.493 0.505	0.211 0.359	1.013 1.111	0.737
Jun 1994	49	0.463 0.528	0.375 0.419	0.524 0.589	-2.264*
Jul 1994	No	data			
Aug 1994	36	0.433 0.464	-0.045 0.275	0.572 0.872	-2.009*
Sep 1994	59	0.437 0.454	0.243 0.389	0.597 0.602	-1.889*
Oct 1994	93	0.578 0.522	0.612 0.462	0.632 0.525	-1.433
Nov 1994	158	0.509 0.576	0.623 0.650	0.604 0.662	0.754
Dec 1994	185	0.484 0.511	0.482 0.527	0.804 0.816	0.213
Jan 1995	73	0.533 0.561	0.444 0.534	0.937 0.974	1.588

anced by an improvement due to averaging out the random errors.

The model skills, for each of the four measures at day 6, per month from April 1993 through January 1995, is given in Table 2. The number of buoy-days available for verification is given, then each of the three skill measures (Skiles model then new model), and finally the *t* statistic for the error radius. Again, there is little difference in character between the scoring methods. All are nonseasonal, though the vector correlation and linear correlation have substantial scatter. Often the error radius shows no statistically significant (at 95% level) difference between the models. Of the 10 months when the difference is significant, the new model is better in 9. The magnitude of the difference in mean error radius is only a few tenths of a kilometer in most months, versus magnitudes of drifts that are on the order of 40 km.

4. Conclusions

We have shown that the revised virtual floe model, based on the Thorndike and Colony (1982) model, is superior to the Skiles (1968) drift law forecast model. We find no dependence of skill, by any measure, on forecast length out to day 6. All four measures of skill give essentially the same impression of model performance, so that ice modelers may continue to use whichever measure they prefer. Skill scores given previously (Thorndike and Colony 1982; Flato and Hibler 1992) are higher than those seen here, very likely because this model is working with forecast rather than analyzed

fields. This model was implemented in operations in October 1997. It is available on the World Wide Web at ftp://polar.wwb.noaa.gov/ice/drift.out and http://polar.wwb.noaa.gov/seaice/, and via GTS under headers FZXX41 KWNO (Alaska subregion output) and FZAK41 KWNO (global).

Acknowledgments. I would like to thank D. B. Rao, L. Burroughs, V. M. Haliburton, A. G. Haliburton, J. Waldrop, G. Flato, B. Colman, and an anonymous reviewer for their editorial assistance. My thanks also to C. Sercy, NWS Alaska Region, for the software for producing the Web graphics. Thanks also to L. Burroughs for his help with the figures and implementation, and to L. Breaker for the code and explanations of vector correlation.

REFERENCES

Brier, G. W., and R. A. Allen, 1950: Verification of weather forecasts. *Compendium of Meteorology*, Amer. Meteor. Soc., 841-848.

Crosby, D. S., L. C. Breaker, and W. H. Gemmill, 1993: A proposed definition for vector correlation in geophysics: Theory and application. *J. Atmos. Oceanic Technol.*, **10**, 355-367.

Flato, G. M., and W. D. Hibler III, 1992: Modeling pack ice as a cavitating fluid. *J. Phys. Oceanogr.*, **22**, 626-651.

Ip, C. F., W. D. Hibler III, and G. M. Flato, 1991: On the effect of rheology on seasonal sea-ice simulations. *Ann. Glaciol.*, **15**, 17-25.

Martinson, D. G., and C. Wamsler, 1990: Ice drift and momentum exchange in winter Antarctic pack ice. *J. Geophys. Res.*, **95**, 1741-1755.

Nansen, F., 1902: *The Oceanography of the North Polar Basin: The Norwegian North Polar Expedition 1893-1896*. Scientific Results, Vol. 3, 427 pp.

- Preller, R. H., and P. G. Posey, 1989: The Polar Ice Prediction System—A sea ice forecasting system. NORDA Rep. 212, Code PDW 106-8, Washington, DC, 45 pp. [Available from NORDA, Washington, DC 20361.]
- Skiles, F. L., 1968: Empirical wind drift of sea ice. *Arctic Drifting Stations*, Arctic Institute of North America, 239–252.
- Thorndike, A. S., and R. Colony, 1982: Sea ice motion in response to geostrophic winds. *J. Geophys. Res.*, **87**, 5845–5852.
- Willmott, C. J., S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Donnell, and C. M. Rowe, 1985: Statistics for the evaluation and comparison of models. *J. Geophys. Res.*, **90**, 8995–9005.