Title: Methodology and Architecture for Using Oceanographic and Meteorological Ensembles to Improve Forecast Skill and Guidance in Support of Tactical Planning and Operations

Authors:

James Etro, Itri Corporation, 7208 Hadlow Drive, Springfield, VA 22152
jim.etro@itricorp.com

Dr. James Hansen, Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 93943
jim.hansen@nrlmry.navy.mil

John Cook, Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 93943
john.cook@nrlmry.navy.mil

Arthur Etro, Itri Corporation, 7208 Hadlow Drive, Springfield, VA 22152
arthur.etro@itricorp.com

ABSTRACT

Ensembles of oceanographic and meteorological numerical forecast models are an excellent tool for quantifying the uncertainties in the marine environment that impact tactical operations. One challenge in exposing the information rich ensemble data to military operators and to existing Tactical Decision Aids (TDAs) is to distill the information into a look and feel that the operator quickly understands and to format it so that it can be put into TDAs. Our approach is to use the ensemble to improve forecast skill and guidance as it applies to the specific operation being supported without exposing the consumer to the probabilistic information about the ensemble. In a post processing engine, known as the Ensemble Forecast Application System (EFAS), the ensemble is distilled into a forecast that has a familiar deterministic look and feel for the operator, and it fits into their existing TDAs. EFAS first applies a bias-correction to some of the ensemble parameters to improve their forecast skill. For example; we see that forecasts of temperature, pressure, wave height, and wind speed are improved by applying bias-corrections, while wind direction forecasts are not. Then by using a consensus finding algorithm based on the RMSE history of the forecast parameter, it is possible to select the most skillful forecast value or forecast field or member from the ensemble. From the ensemble members, one can also extract a spread around that forecast value by forming the probability density function from the ensemble members. The parameter's probability density function, based on the operator's requirement for accuracy or the operator's tolerance in the forecast error, can also be used to state the confidence (high or low) that the forecast will meet the operator's needs as it pertains to the operation. The architecture for such a forecast system can be implemented in three virtual locations. The ensemble generation is accomplished at the large central computing sites. The on-scene operator interfaces with the post processing engine to derive the specific information needed to support their operation. And the post processing engine which aggregates and distills the ensembles for use, can be at any location deemed appropriate by the area commander, provided the engine has access to the raw ensemble data sets.

Keywords: METOC Ensembles; Consensus Forecasts; Bias-Correction; Forecast Products; Decision Aids; Forecast Skill, Ensemble RMSE

1. Introduction

> Know the enemy, know yourself; your victory will never be endangered.
> Know the ground, know the weather; your victory will then be total[1].

In planning and during the execution of military operations the physical and intellectual resources needed to know enemy, self, ground, and weather are large and complex; and they are essential elements in developing plans and making good decisions. Planners and operators must factor in a several classes of time variant information and take action on short notice as the operation unfolds. Some of these information types include friendly and enemy orders of battle, weather and sea state, and logistics. To plan and act effectively they must process a large amount of information quickly and they can only use what is immediately at hand and easily understood and trusted in the context of the operation.

As one of the essential elements of a good decision, an accurate forecast of weather and ocean conditions, and an ability to convey a measure of confidence in the forecast, is critical. Meteorological and Oceanographic (METOC) ensembles can be an important source of this forecast information; but, METOC ensembles are voluminous, provide ambiguous and redundant information, are difficult to quickly and easily interpret, and are even more difficult to apply directly to operational decisions. Thus, the value that METOC ensemble prediction brings to problem solving will be overlooked if the information isn't in a form that makes it immediately understandable and actionable. Decision makers and operators require weather and ocean analyses and forecasts that are accurate, in a form that they can quickly understand and put into their Tactical Decision Aids (TDAs), and they must know that the information is reliable. With respect to METOC analyses and forecasts this means:

- Having easy access to the best data and information pertinent to their operation.
- Having an assessment of the degree of confidence in the forecast.
- Interfacing applications (forecast products and TDAs) that translate the data and information into an actionable characterization of the environment for the operator.

Our goal in developing the Ensemble Forecast Application System (EFAS) is to derive a method for allowing operations personnel to exploit METOC ensemble information in their time sensitive planning and execution processes. Our approach is to distill the METOC ensemble information into an accurate representation of the needed information that can be quickly/simply injected into Forecast Products and the TDAs that are routinely used, and to provide a simple measure of the confidence in the accuracy of the METOC information. In achieving our results we adhered to the following guidelines:

- The system developed would use ensemble data from any METOC ensemble produced by a national level central-site computing facility. For example; the system would be capable of using ensembles from the US Navy Fleet Numerical Meteorology and Oceanography Center (FNMOC), US Naval Oceanographic Office (NAVOCEANO), NOAA National Centers for Environmental Prediction (NCEP) Central Operations, European Centre for Medium-Range Forecasts (ECMWF), and all other national centers that output their ensemble model data in GRIB and/or any other standard data format.

[1] Sun Tzu, *The Art of War*, c.400-320 BC

- Results or output would be targeted and formatted for non-scientific personnel and they would be simple to understand and easy use in operational planning and in existing TDAs. This implied that we would not expose the operator the forecast products or the TDAs to typical probabilistic data output and formats but strive to maintain a deterministic "look and feel" that leverages existing TDA capabilities.
- Results would be available in an operationally feasible timeframe. This led us to implement short post processing training periods and simple bias-correction and spread adjustment techniques.

In our use of the ensemble data sets the highest priority was focused on improving the forecast skill of a deterministic mid-range (24-144hr) forecast. Our approach was to distill the information from the ensemble members to form a more accurate consensus forecast that could be put into existing forecast product and TDAs. Although the statistical spread of possible forecast values, including the possibility of an extreme event, may be a consideration when forecasters examine ensembles for guidance, we are examining the members to establish a consensus forecast and to assess our confidence in that specific forecast. Our technique focuses on improving a deterministic forecast rather than ensuring that all possible forecasts are exposed to the operational decision maker.

2. Ensemble Post Processing

The EFAS post-processing calibration is initiated after the ensemble GRIB files (or segments of files) have been downloaded from the production facility. EFAS processing is controlled by a configuration file specifying the forecast parameters and data levels (e.g. sea level pressure, 10 meter wind speed and direction, 500mb temperature, significant wave height, total sky cover, etc.) required from the numerical weather and/or ocean model ensembles. EFAS organizes the selected fields into its data base and computes a bias-correction to the parameter across every ensemble member for each grid point. Once bias-corrected, the members can be ready for spread adjustment techniques. In our work to date, we do not apply spread adjustments because we haven't identified a spread adjustment algorithm that shows improved forecast skill. After the calibration, the "consensus forecast" is established and the forecast information extracted for use by a forecaster and their decision aids. The calibrated ensemble is also used to establish a probability distribution function of the desired forecast parameter that is used to compute a confidence measure describing the accuracy of the consensus forecast within the user's specified tolerance for error.

The flow of the EFAS process (from left to right) as it is implemented today is depicted in Figure 1.

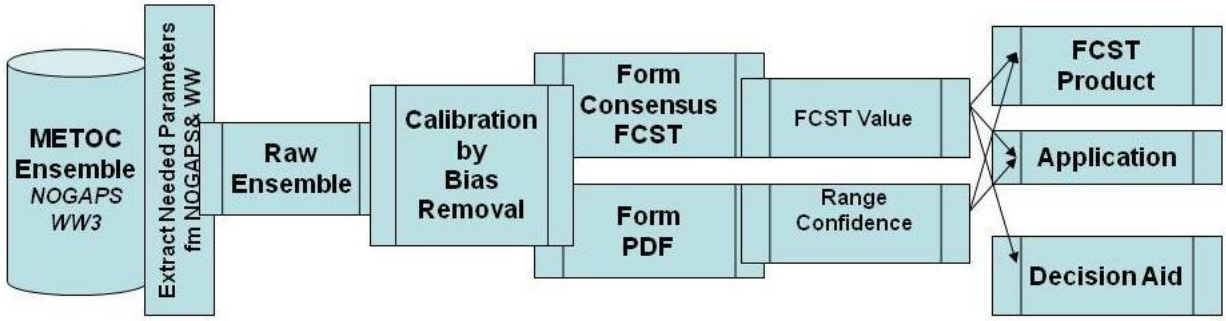

Figure 1. Task Flow

Bias-corrections

The METOC ensemble that is down loaded from a central site or from multiple sites[2] and organized in the EFAS data base is referred to as the Raw Ensemble. The ensemble data calibration process is initiated by applying a bias-correction to every grid point and level, and at every forecast step (also known as forecast TAU) that a result is needed for. In the cases addressed in this paper, operational METOC ensemble data from the FNMOC Navy Operational Global Atmospheric Prediction System (NOGAPS) and WAVE WATCH III (WW3) models are used in a marine forecast and ship routing algorithm. Bias-correction was applied to the needed forecast parameters at 6 hour intervals across the full set of available forecasts out to TAU 144hrs. For each for forecast TAU (6hr, 12hr, 18hr… 144hr), at every grid point each of the ensemble members' forecast parameter value was subtracted from the verifying analysis value (also known as the Analysis or TAU 0). The average difference across the ensemble members between the forecast and analysis at each grid point and TAU into the future comprise the bias value at that grid point. For each forecast cycle (00Z forecast cycle and 12Z forecast cycle) we then compute a running mean of the last 30 days of 00Z and then 12Z forecast bias values at each TAU and grid point and apply the new 30-day running mean grid point bias-correction to each ensemble member). This data set is referred to as the Bias-corrected Ensemble. This process is repeated at every 12 hour forecast cycle after the raw ensemble is produced.

The equation applied at every grid point is:

B = (1/M)(1/N){ (Sum of 1 to M over m) of (Sum(1 to N over d) of [P(i,j,k,t,m,d) - Pobs(i,j,k,t,d)]}

- i,j,k,t originates at time = 0, longitude 0°, and latitude 90° north
- B is the bias-correction for a given forecast parameter at location i,j,k and TAU
- P*(i,j,k,t,m, d=0) = P(i,j,k,t,m,d = 0) - B
- Where P* is the unbiased forecast for each grid point (i,j,k), each TAU (t), and each ensemble member (m)
- P(i,j,k,t,m,dd) is the raw forecast parameter for each point, TAU, and ensemble member on any given model run (d).
- N is the number of model runs (d) of history (i.e. if 30 days of 12Zforecasts; d = 30) stored for correcting the bias.
- d = 0 is the current forecast time.
- M is the number of ensemble members (m).

In our use of the ensemble data sets the highest priority was to improve the mid-range (24-144hr) forecast skill. Accordingly, the number of forecast cycles used to establish the bias-correction becomes an important consideration. If the averaging period is too short or too long, the bias-correction will not capture the relevant model bias. It is our experience that the bias-correction averaging period should be long enough to recognize the synoptic scale errors, but short enough to prevent seasonal biases from over influencing or smoothing out the synoptic bias-correction. We tested periods from 05 to 40 days and found that for our study using a 24-30 day period provides the best results (an accurate forecast).

Another consideration in keeping the averaging period short is that if a new forecast parameter is required it can be brought 'up-to-speed' as soon as possible and if a new mesoscale ensemble domain is set-up it can be used as quickly as possible. In the future we intend to test replacing the bias-correction running

[2] EFAS is capable of post processing multi-model ensembles.

average algorithm with a Kalman filter. A Kalman filter may reduce the averaging period from 30 days to 5 days or less (Delle Monache et al. 2008).

Table 1 summarizes a three month test of the NOGAPS and WW3 bias correction method described above. The table shows the summary result of comparing raw and bias-corrected forecasts against their verifying analysis and also against a randomly selected single member forecast (labeled Single Model, simulating a deterministic forecast) and the ensemble average of the raw and bias-corrected ensemble forecasts. In all cases that we examined (out to 48 hr forecasts) over the three month period the ensemble averages produced the superior results compared to the deterministic forecasts (a random single ensemble member). The impact of this finding is that a forecaster can rely on the ensemble average for guidance and the results can be confidently provided to decision aids that require deterministic inputs. But, it is important to point out that although the ensemble average is an excellent predictor, in every ensemble there is at least one member which is a better forecast. We plan to test whether it is possible to exploit our statistics to guide the selection of the most likely member to provide the most accurate forecast. The advantage of the "most likely" member forecast over the ensemble mean forecast is the quality of the gradients and the physical consistency of the deterministic forecast. As the ensemble solutions diverge over the length of the forecast, the averaging process destroys the gradients and creates unrealistically smooth fields. And since the ensemble averaging process is applied at each level and for each parameter independently, the physical relationships are not preserved creating unrealistic vertical gradients and inconsistencies between parameters.

Parameter	**Tau**	**Best Ensemble Average Conditioning Method Compared to Analysis**	**Differences Compared to Analysis**		
			Average Difference		**Improvement**
			Analysis – Single Model	**Analysis – Best Ensemble Mean**	
Sig Wave Hgt (meters)	6	Bias-correction	0.22	0.16	.058 m
	24	Bias-correction	0.24	0.17	.069 m
	48	Bias-correction	0.31	0.23	.083 m
SFC Wind Speed (knots)	6	Bias-correction	1.04	0.64	.4 m/s
	24	Bias-correction	1.38	1	.38 m/s
	48	Bias-correction	1.68	1.25	.43 m/s
SFC Wind Direction (deg)	6	Raw Ensemble	30	23	7°
	24	Raw Ensemble	38	30	8°
	48	Raw Ensemble	47	38	9°

Table 1. Forecast Differences

3. Formulation of the Consensus Forecast

We consider the Consensus Forecast of a parameter the 'most accurate' forecast that the ensemble can provide. For the parameters we examined (surface pressure, near surface wind and air temperature, sea swell, wind wave, and total cloud cover) the ensemble average produces the best forecast. Figure 2 shows a set of representative forecasts of the ensemble average of sea level pressure from the raw ensemble and the bias-corrected ensemble. Also shown is the verifying sea level pressure analysis. The ensemble averages very closely resemble the analyzed synoptic meteorological pattern giving us confidence that the forecasts derived from ensemble averages is accurate.

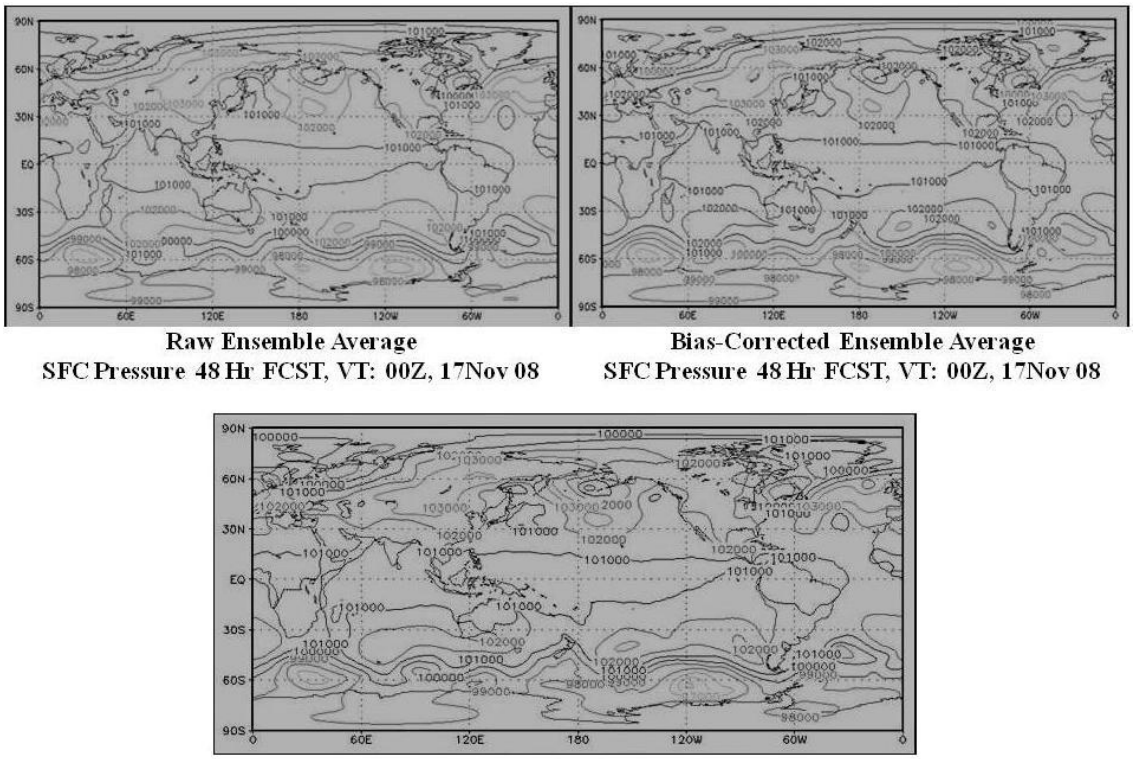

Figure 2. Averaged Raw and Bias-Corrected Ensembles with Verifying Forecast

For each parameter and forecast TAU a consensus forecast is then selected from either the raw ensemble or the bias-corrected ensemble averages. The forecast from the raw or the bias-corrected ensemble is selected by examining the root mean square error (RMSE) of the raw ensemble and the bias-corrected ensemble averages differenced against the verifying analyses of each parameter. The forecast value is then selected from the more accurate forecast (smallest difference) between the two.. Table 2 summarizes the RMSE between the raw ensembles and the bias-corrected ensembles predictors of swell height, swell direction, swell period, 10 meter wind speed, and 10 meter wind direction at 24 hour intervals to 144 hours, over the entire globe, during the period 30 March 2010 through 11 November 2010. Table 3 summarizes the RMSE between the raw ensembles and the bias-corrected ensembles predictors of 10 meter wind speed and 10 meter wind direction at TAU 72 hrs, in three separate latitude zones (30°S - 90°S, 30°N - 90°N, and Global), during the period 1 June 2010 through 31 August 2010. From the type of analysis demonstrated in tables 2 and 3, if we need a global forecast field we see that for sea swell and wind wave forecasts to 144 hours, the bias-corrected ensemble average generally provides a more accurate forecast than the raw ensemble average for all parameters. However, if we need a wind forecast in the North Atlantic during the summer season we see that the 10 meter wind speed is best predicted by the bias-corrected average and that the 10 meter wind direction is best predicted by the raw ensemble average. In the future we intend to examine the RMSE of the predictors in smaller regional locations such as areas near and downstream of western boundary currents and near and downstream of large continental mountain ranges to see if the same types of relationships exist.

Parameter	Tau	Raw Ensemble RSME	Bias-corr Ensemble RSME
swl_hgt (m)	24	0.3647	0.31
	48	0.42	0.36
	72	0.48	0.44
	96	0.56	0.52
	120	0.64	0.61
	144	0.71	0.69
swl_wav_dir (deg)	24	15.0862	14.7914
	48	17.7703	17.5793
	72	21.3348	21.2089
	96	25.4258	25.3554
	120	29.7426	29.7280
	144	33.9047	33.9248
swl_wav_per (sec)	24	0.6548	0.5337
	48	0.7152	0.6038
	72	0.7964	0.6919
	96	0.8919	0.7931
	120	0.9942	0.9000
	144	1.0970	1.0059

Parameter	Tau	Raw Ensemble RSME	Bias-corr Ensemble RSME
wind_spd (m/s)	24	1.43	1.38
	48	1.77	1.73
	72	2.08	2.04
	96	2.33	2.30
	120	2.54	2.51
	144	2.69	2.68
wind_dir (deg)	24	28.04	27.80
	49	35.30	35.14
	72	42.52	42.43
	96	49.33	49.27
	120	55.26	55.19
	144	60.18	60.10

The summary values are the average of every grid point on the 1° global grid output from WW3 and NOGAPS ensembles produced at FNMOC.

Table 2. Parameter RMSE, Global – Averaged over period 30 March 2010 – 11 November 2010

Parameter	Latitude Zone	Tau	Raw Ensemble RSME	Bias-corr RSME
wind_spd (m/s)	30°S - 90°S	72	2.87	2.85
wind_ spd (m/s)	30°N - 90°N	72	2.06	2.01
wind_ spd (m/s)	Global	72	2.07	2.02
wind_dir (deg)	30°S - 90°S	72	42.50	42.45
wind_dir (deg)	30°N - 90°N	72	58.86	58.90
wind_dir (deg)	Global	72	44.05	43.90

Table 3. Parameter RMSE, Zones – Averaged over period 1 June 2010 – 31 August 2010

4. Preparing Raw and Bias-Corrected Ensembles for Use in Forecast Product and Decision Aids

The next step in EFAS is to assemble the consensus forecast results of the parameters needed. As determined from an analysis such as is summarized in Table 3, if a wind speed or wave height forecast is needed we extract the information from the bias-corrected ensemble average. If a wind direction forecast is needed we extract the information from the raw ensemble average. When these individual consensus forecasts are combined we refer to that as the Hybrid forecast or the Hybrid Ensemble. This hybrid combination of forecast values can then be interfaced to a forecast product or decision aid in the same manner as provided by a deterministic forecast.

When using hybrid consensus forecasts, whose source varies by parameter and forecast TAU, it is important to remember that the resulting forecasts are not physically/meteorologically consistent because it is a statistical result. In spite of this contradiction, as demonstrated in Tables 1 and 2, we have found that this method provides more accurate results than if the forecast came from a random member of the ensemble serving as the deterministic forecast. Where physical/meteorological consistency is required, such as a sounding, then this method may not be appropriate for input to decision aids.

As demonstrated in Figure 3, another issue with ensemble averages is that although the ensemble average will regularly provide a superior forecast of a parameter's value over the single deterministic physics-based root numerical model; as the forecast time increases the natural gradients predicted by the ensemble average decrease unrealistically.

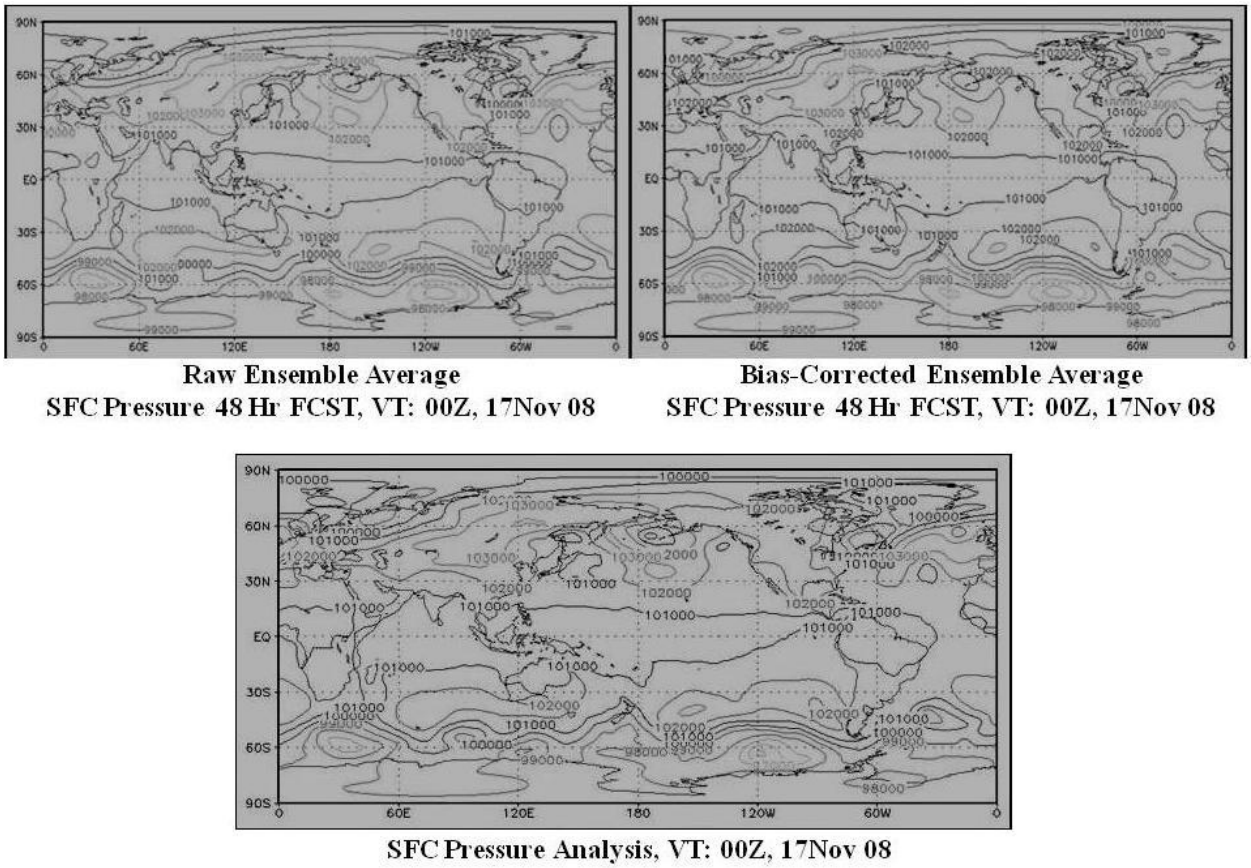

Figure 3. Relaxing Gradient Resulting From Ensemble Member Averaging

When the processes are complete and the consensus forecast is established, we also produce a probability density function (PDF) from the member values to determine a likely range of values around the forecast parameter at every grid point and TAU. In constructing the PDF we use a simple density smoothing of the ensemble members. Using the PDF it is possible to establish a "most likely" forecast value and a range of most likely forecast values (see Figure 4). The Most Likely Value (MLV) data extracted from the raw and the bias-corrected ensembles were evaluated as a potential candidate to the consensus forecast. Our results (not shown) demonstrated that although the MLV was more accurate than the deterministic forecast, it does not provide a better forecast than the ensemble average. However, the PDF does provide additional information about the range of likely values and the range information can

improve the usability of the forecast products. EFAS allows a user-defined spread of probability around the MLV to determine the range.

Note that Figure 4 depicts a tri-modal distribution of the probable forecasts. Our convention in selecting a range applied to a forecast product is to find the forecast value (from Figure 4 we may select the MLV or an average of the possible forecasts) and then bound or range the forecast values from the lower and upper limit of the values which are with-in 20% of the maximum probability of the MLV (highest probability value)

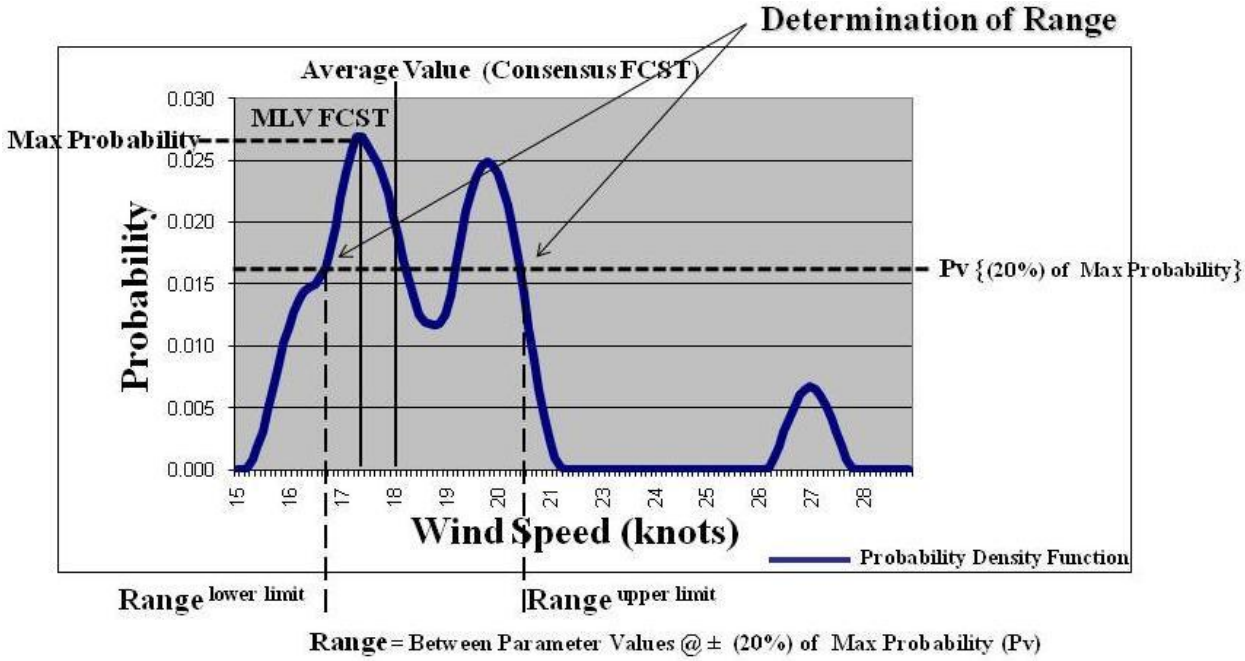

Figure 4. Determination of the Ensembles Most Likely Value (MLV) and Range at a Grid Point

After the consensus forecast is determined it can be interfaced into the TDAs and used to create forecast products. Table 4 demonstrates a sample marine forecast product produced using the consensus forecast and the likely range of forecast values around the consensus that are based on the shape of the PDF and a user specified probability spread.

5. Forecast Confidence

The example marine forecast product in Table 4 introduces another significant benefit of using ensemble information for forecaster guidance and in forecast product: Forecast Confidence. In deriving this term 'Forecast Confidence' our motivation is to deliver information that an operational forecaster may provide to their clients about their confidence in the forecast. The primary consideration in determining the confidence is to establish the purpose of the forecast and then to establish the client's sensitivity or tolerance to an error in the forecast. We define the sensitivity or tolerance of forecast error as the amount of variability that can experienced during the operation before the impact of that parameter on events is important to the forecast consumer. For example; if the forecast is supporting small boat operations an error larger than -1 ft and +2ft may be significant to the planner and operator, but if the forecast is

supporting very large ship transit operations an error larger than +/- 6ft may become the point at which the forecast error is significant.

[Location 50°N, 030°W] [48 Hour FCST Valid at 15 Dec 00Z]			
Parameter	**MLV FCST**	**Range**	**Forecast Confidence**
Wind Speed **Wind Direction**	**14.0 m/s** **104°**	**13 m/s – 14.5 m/s** **101° - 106°**	**High** **Low**
Sig Wave Height	**5.5 m**	**5.0 m - 6.0 m**	**Medium**
Pressure (Altimeter Setting)	**1020 mb**	**1018 mb – 1022 mb**	**High**
Temperature	**13.5 °C**	**13.0 °C – 14.0 °C**	**High**

Table 4. Sample Marine Forecast Product at Geographic Location

When we use the term Confidence in the forecast value of a parameter at a point (x,y,z,t), we mean that the forecast value is within the sensitivity or tolerance of that specific forecast's error. High Confidence means "I'm reasonably certain that the forecasted parameter will be within the tolerance of forecast error because a very large majority of the ensemble members were close enough to the forecast value" and Low Confidence means "I believe it is likely that the forecasted parameter will be well outside the tolerance of forecast error because a very large majority of the ensemble members were outside the tolerance range."

An example of applying Confidence to a specific forecast is for a parameter needed for a marine forecast (Table 4) and the ship routing example provided in Section 7, the tolerance for swell height forecast error is +/- 2 ft. Accordingly for ship routing forecasts we would have HIGH CONFIDENCE in the forecast if ≥ 80% of the ensemble members are all within 2 feet or .61 meters of the value derived from the ensemble to represent the forecast; MEDIUM CONFIDENCE in the forecast if ≥ 50% and < 80% of the ensemble members are all within 2 feet or .61 meters of the value derived from the ensemble to represent the forecast; and LOW CONFIDENCE in the forecast if < 50% of the ensemble members are all within 2 feet or .61 meters [configurable] of the Value derived from the ensemble to represent the forecast. The confidence limits and threshold values are configurable in EFAS.

In our example of a marine forecast (Table 4) and the ship routing example, the confidence in the wind forecast is derived from a wind speed tolerance of +/- 5 knots in the speed forecast and +/- 10° in the direction forecast. In the wind speed forecast HIGH CONFIDENCE in the wind speed forecast is when ≥ 80% of the ensemble members are all within 5 knots or 2.57 m/s of the Value; MEDIUM CONFIDENCE is when ≥ 50% and < 80% of the ensemble members are all within 5 knots or 2.57 m/s of the value; and LOW CONFIDENCE is when < 50% of the ensemble members are all within 5 knots or 2.57 m/s of the MLV.

6. Applying METOC Ensembles to a Ship Routing Forecast and Algorithm (or TDA) Applications

Once the ensembles have been post processed, the specific consensus forecast information supporting an application is extracted and interfaced or input to the forecast product or a TDA. In some cases it is useful to expose the 'operator' to the forecast product (value, range, confidence) as demonstrated in the Marine

Forecast Product (Table 4). In other cases it may be desirable to provide the decision aid the consensus forecast noted as the 'best forecast'. There may also be cases where it is desirable to interface all of the ensemble members to the TDA so that one can review the full set (or an ensemble) of decision aid results. Figure 5 demonstrates an ensemble of ship routes optimized for minimum fuel burn and to avoid high winds and seas. The ensembles of routes were computed from a 35 member METOC ensemble (16 raw ensemble members, 16 bias-corrected members, 1 raw ensemble average, 1 bias-corrected average, and 1 hybrid average) of swell direction, swell height, swell period, 10 meter wind speed, 10 meter wind direction, wind wave direction, wind wave height, and wind wave period.

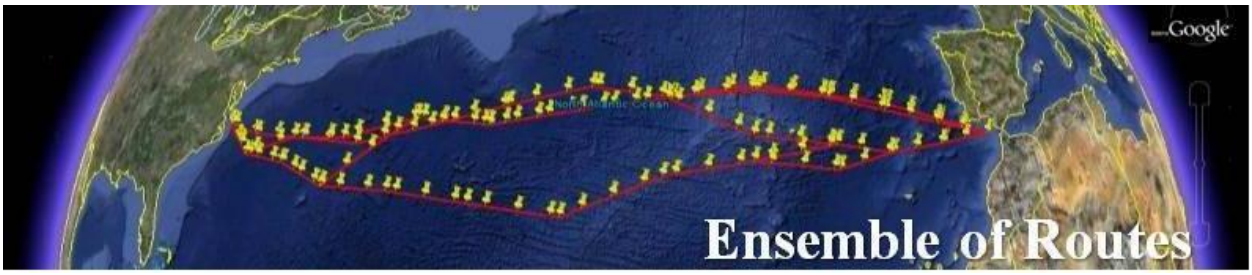

Figure 5. Ensemble of Ship Routes (Gibraltar to Norfolk, VA)

Note that when all of the 35 routes resulting from the 35 ensemble members are plotted many of them over plot each other. This is a function of the sensitivity of the optimum track routing algorithm to the wind and sea inputs and their influence on the hull resistance and the wave and wind avoidance algorithms. Even though there are many over plots there are clearly several optimum routes that are predicted or recommended. As a decision aid this view of the available information isn't very satisfying. Which is the best route to recommend that the mariner should sail?

To answer that question the route planner can benefit from using the consensus forecast. In the case of the ship routing algorithm our consensus (hybrid) forecast is derived from interfacing to the parameters in the table:

Hybrid Forecast for Routing Application	6 Hr – 144 Hr Forecast	> 144 Hr
SWL_WAVE_DIR	BIAS_CORR	Climatology
SWL_WAVE_HT	BIAS_CORR	Climatology
SWL_WAVE_PER	BIAS_CORR	Climatology
WIND_DIR	RAW	Climatology
WIND_SPD	BIAS_CORR	Climatology
WIND_WAVE_DIR	BIAS_CORR	Climatology
WIND_WAVE_HT	BIAS_CORR	Climatology
WIND_WAVE_PER	BIAS_CORR	Climatology
Surface Currents	Root Forecast (no Ensemble Available)	Climatology

Table 5. Parameters that Comprise the Hybrid Ensemble

These are considered the most accurate representation of the forecast that is available to the routing algorithm. Note that from the analysis described in section 3, the bias-corrected average ensemble forecast is used for the parameter inputs except the raw ensemble average is used for the wind direction forecast. When we test for accuracy of the ensemble average wind direction forecasts in the northern hemisphere, we see that the raw ensemble average provides the best forecast.

The route forecast resulting from the Hybrid (Consensus) Forecast is depicted in Figure 6. Because this route is derived from the 'most accurate' weather forecasts over the entire voyage, it is also considered the best route recommendation.

Figure 6, Hybrid or Consensus Route Forecast

7. Future Work Planned

There are several improvements or upgrades that are important to EFAS. The following topics are those that we intend to pursue in the near future; they are listed in order of increasing complexity.

a. Improvement to the Bias-correction Technique

In the near term we intend to improve on the Bias-correction technique by replacing the 30-day running average Bias-correction technique with a Kalman filter. Once the Kalman filter is installed in EFAS we will be able to check the algorithm's skill for improving the ensemble members as predictors of forecast values. If the filter provides a consistently more accurate forecast we will replace the 30-day running average algorithm with the Kalman filter as our Bias-correction technique and derive our consensus forecasts from it. It will be interesting to see how the Kalman filter handles the parameters with a directional component that the 30-day running average didn't handle well.

b. Merge a mesoscale model's ensemble into the global ensemble

The EFAS architecture easily supports the introduction of multiple ensembles because all of the calibrations or calculations are accomplished at every grid point, level, and TAU independently. Presently we are using the ensembles from a global model such as NOGAPS on a 1° grid. In an independent computing environment we also experimented on a mesoscale model, the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS®). The COAMPS ensemble was a 16 member ensemble with a nest of 3 grids in the computational domain. Excellent results were obtained from this limited nested grid ensemble so we think that it will be beneficial to nest a mesoscale ensemble into a coarser global model to achieve more accurate mid-range forecasts in locations where considerations of mesoscale impacts may produce higher fidelity forecasts.

c. Develop a technique for using the Most Likely Value to select the best ensemble member to be the forecast.

Although the average of the ensemble members consistently produces the best (most accurate) forecast at each grid point, the full field forecast is not meteorologically consistent because each grid point and level are post processed independently. Since we know that there is often an ensemble member that will make a better forecast than the average of the members it may be possible to statistically determine that member. Based from our experience in distilling a Most Likely Value from the ensemble PDF, we intend to pursue a method for determining the Most Likely Forecast Member of the METOC ensemble so that we have horizontally and vertically consistent meteorological fields that provide the best consensus forecast.

Acknowledgements
The EFAS project was sponsored by the Office of Naval Research (ONR) under the Small Business Technology Transfer (STTR) Contract N00014-08-C-0173. Approved for public release; distribution is unlimited (NRL 10-1226-4806).

References

DELLE MONACHE, L., WILCZAK, J., MCKEEN, S., GRELL, G., PAGOWSKI, M., PECKHAM, S., STULL, R., MCHENRY, J. and MCQUEEN, J. (2008), A Kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone. Tellus B, 60: 238–249. doi: 10.1111/j.1600-0889.2007.00332.x