Reanalysis of the PacIOOS Hawaiian Island Ocean Forecast System, an implementation of the Regional Ocean Modeling System v3.6

Dale Partridge[1] **and Brian S. Powell**[1]

[1] University of Hawaii at Manoa, Department of Oceanography, Marine Sciences Building, 1000 Pope Road, Honolulu,

Hawaii 96822, USA.

Geoscientific
Model Development
Discussions

Open Access

Abstract

A 10-year reanalysis of the PacIOOS Hawaiian Island Ocean Forecast System was produced

using an incremental strong constraint 4D-Variational data assimilation with the Regional Ocean

Modeling System (ROMS v3.6). Observations were assimilated from a range of sources: satellite-

derived sea surface temperature (SST), salinity (SSS), and height anomalies (SSHA); depth

profiles of temperature and salinity from Argo floats, autonomous SeaGliders, shipboard conductivity-

temperature-depth (CTDs); and surface HFR velocity measurements from high frequency radar

(HFR). The performance of the state-estimate is examined against a free-running forecast show-

ing an improved representation of the observations, especially the realization of HFR surface

currents. EOFs of the increments made during the assimilation to the initial conditions and

atmospheric forcing components are computed, revealing the variables that are influential in

producing the state-estimate solution and the spatial structure the increments form.

1 Introduction

The Pacific Integrated Ocean Observing System [*PacIOOS*, 2018] has produced daily

forecasts of the ocean state surrounding the Hawaiian Islands since 2009. To facilitate the fore-

casts a data assimilation procedure is used to incorporate recent observational data into the model

to produce the optimal initial state from which to forecast. A number of modeling studies have

been performed with older versions of this model to examine various features of the model-

ing framework, such as the state estimation [*Matthews et al.*, 2012], nested models [*Janeković

et al.*, 2013] and the vorticity budget [*Souza et al.*, 2015]. In this work, we perform an extended

reanalysis from 2007 to 2017 to coincide with the upgrade of the forecast system to a newer

model version and improved processing, in order to produce a consistent data set for further

study of the dynamics around Hawai'i.

The PacIOOS forecast system uses the time-dependent Incremental Strong constraint 4-

dimensional Variational Data Assimilation (I4D-Var) scheme [*Courtier et al.*, 1994; *Moore et al.*,

2004] within the Regional Ocean Modeling System (ROMS) [*Moore et al.*, 2011a; *Powell et al.*,

2008; *Matthews et al.*, 2012] to best reduce the residuals between the model and observations,

while maintaining a physically consistent solution. The class of methods known as 4D-Var are

state-estimation techniques that create a quadratic cost function to be minimized over a de-

fined time window, utilizing observations at the time they occur in a physically consistent man-

ner to adjust the initial state, boundary conditions, and atmospheric forcing to represent the

measurements. The I4D-Var scheme is used in operational centers around the world and solves

for increments to the model state, boundary conditions, and atmospheric forcing using the model physics as a constraint. The combination of I4D-Var within ROMS has been used in previous studies of various regions [*Powell et al.*, 2008; *Broquet et al.*, 2009; *Zhang et al.*, 2010; *Matthews et al.*, 2012; *Souza et al.*, 2015]. The details of the model and the observations used within this study are provided in Section 2.

Our model domain covers the Hawaiian Island Archipelago (Figure 1), a dynamically active region for both the ocean and atmosphere. The North Equatorial Current (NEC), flowing from the east, splits upon encountering the island of Hawai'i, with the bulk transport traveling around the south of the island and continuing west, while the North Hawaiian Ridge Current (NHRC) follows the ridge of the other islands in the chain to the north. In the atmosphere, there are persistent trade winds from the northeast that, combined with steep mountainous terrain on the islands, cause wind wakes in lee of the peaks, particularly on the islands of Hawai'i and Maui. This introduces strong temperature gradients, increases the seasonal variability [*Sasaki and Klein*, 2012], and drives smaller currents such as the Hawaiian Lee Countercurrent (HLCC) [*Smith and Grubišić*, 1993; *Xie et al.*, 2001; *Chavanne et al.*, 2002].

There are two main objectives to this study: to assess the skill and performance of the state-estimation model, and to analyze the increments made to the initial, boundary and atmospheric forcing terms. For the first objective, we compare the state-estimate solution with a free-running forecast over the decadal time period and examine how the performance changes over time, utilizing observations derived from satellites and *it situ* measurements. In addition, PacIOOS operates seven high-frequency radar stations sites across the Hawaiian Islands. The first station was constructed in 2010, with the remaining six becoming operational over the period from 2011-2015. These instruments produce high resolution (both spatially and temporally) surface current velocities in the vicinity of the islands of O'ahu and Hawai'i. The use of HFR observations within a state-estimation scheme has been shown to produce a significantly improved representation of surface currents [*Souza et al.*, 2015; *Kerry et al.*, 2016]. The impact of the radar stations will be a key focus point. The performance assessment is achieved through the statistics produced by the state-estimation in Section 3, followed by a comparison with observations in Section 4. The forecast skill, a measure of the accuracy for a forecast system is computed with reference to a persistence assumption.

Section 6 focuses on the second objective of the paper, to examine the increments to the initial state and atmospheric forcing to determine how the model is adjusted. By evaluating the Empirical Orthogonal Functions (EOFs) of these increments we determine the spatial pat-

Geoscientific
Model Development
Discussions
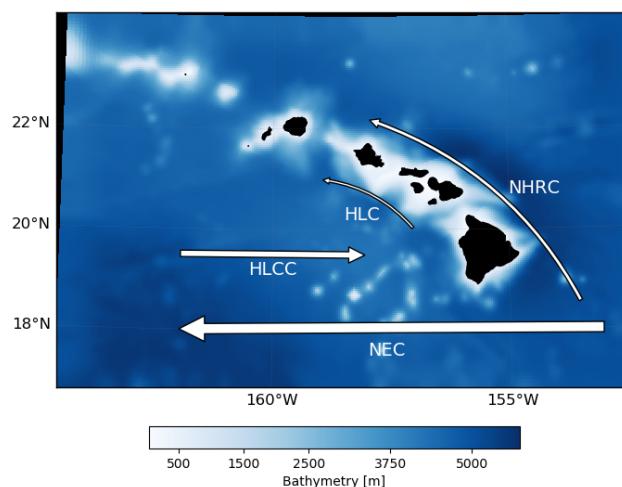
Figure 1. Model domain and bathymetry, with mean currents labelled from *Lumpkin and Flament* [2013].

terns in the variability. Since physical modes are not always independent [*Simmons et al.*, 1983],

the interpretation of EOF modes must be undertaken with some caution. As such the result-

ing modes will not necessarily represent a physical phenomenon, but will highlight the vari-

able spatial patterns made over time by the I4D-Var algorithm.

2 Numerical Model and Data Assimilation System

2.1 Model Configuration

The Regional Ocean Modeling System (ROMS) version 3.6 is used to simulate the phys-

ical ocean around the Hawaiian Islands. ROMS is a free surface, hydrostatic, primitive equa-

tion model using a stretched coordinate system in the vertical to follow the underwater ter-

rain. In order to allow varying time steps for the barotropic and baroclinic components, ROMS

utilizes a split-explicit time stepping scheme (for more details on ROMS, see *Shchepetkin and*

McWilliams [1998, 2003, 2005]).

The Hawaiian Island domain covers 164°W to 153°W longitude and 17°N to 23°N lat-

itude, with bathymetry provided by the Hawaiian Mapping Research Group [*HMRG*, 2017],

shown in Figure 1. The grid has 4km horizontal resolution with 32 vertical s-levels, config-

ured to provide a higher resolution in the more variable upper regions. The configuration model,

Geoscientific

Model Development

Discussions

88   including the method for assimilating surface HFRs and the associated vertical stretching scheme,

89   is identical to the one first presented in *Souza et al.* [2015].

91   Tidal forcing is produced using the OSU Tidal Prediction Software (OTPS) [*Egbert et al.*,

92   1994], which is based on the Laplace tidal equations from TOPEX/Poseidon Global Inverse

93   Solution (TPXO). Tidal constituents included in this simulation are the eight main harmon-

94   ics; M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1, as well as two long period and one non-linear constituent;

95   M_f, M_m and M_4. To avoid any long term drifting of the tidal phases related to constituents we

96   do not consider, the tidal harmonics are updated each year to define the phases in terms of the

97   start of that year.

98   Lateral boundary conditions are taken from the HYbrid Coordinate Ocean Model (HY-

99   COM) [*Chassignet et al.*, 2007] and are applied daily. Within ROMs, we apply the boundary

100   differently for each variable; Chapman [*Chapman*, 1985] conditions are applied to the free sur-

101   face, Flather [*Flather*, 1976] conditions for transferring momentum from 2D barotropic en-

102   ergy out of the domain, while the 3D momentum and tracers variables are clamped to match

103   HYCOM. A sponge layer of 12 grid cells (48km) linearly relaxes the viscosity by a factor of

104   four and diffusivity by a factor of two close to the boundary to account for imbalances between

105   HYCOM and ROMS.

106   From 2007-2009, atmospheric forcing is provided by coarse 2° resolution NCEP reanal-

107   ysis fields. The wind fields are statistically combined with a fine-scale PSU/NCAR mesoscale

108   model (MM5) [*Yang et al.*, 2008a]. From July 2009, atmospheric forcing is provided locally

109   by a high-resolution Weather Regional Forecast (WRF) model [*WRF-ARW*, 2017]. WRF sup-

110   plies information about surface air pressure, surface air temperature, long- and short-wave ra-

111   diation, relative humidity, rain fall rate, and 10m wind speeds. The ocean model is forced us-

112   ing this data every six hours, taken from the atmospheric model with 6km resolution across

113   the entire domain.

114   Prior to the experiment, a six-year non-assimilative model was run using the same ini-

115   tial state, boundary conditions, and atmospheric forcing. The variability of the model is used

116   to produce an estimate of the background error covariances used within I4D-Var, as well as

117   the mean sea surface height to use with sea level anomaly observations.

118   A detailed derivation of the I4D-Var cost function can be found in [*Kerry et al.*, 2016;

119   *Penenko*, 2009; *Weaver et al.*, 2003; *Stammer et al.*, 2002; *Talagrand and Courtier*, 1987]. To

120   formulate the solution, we must provide estimates of the uncertainty in both the model and

121   observations. The model uncertainty, $\mathbf{P}$, is estimated using the variability of the six-year run

122 described above, while observation uncertainty, $\mathbf{R}$, is assumed to be diagonal, (i.e. observa-

123 tions are independent). The implementation of I4D-Var in ROMS is covered extensively in [*Moore*

124 *et al.*, 2011a,b,c].

2.2 Experiment Setup

126 The reanalysis covers a period of 10 years, from July 2007 to July 2017. The minimiza-

127 tion of the cost function first updates the nonlinear model in an *outer* loop, before perform-

128 ing a linear least-squares procedure over multiple *inner* loops. For this grid there is a suffi-

129 cient reduction in J using a single outer loop with 13 inner loops to efficiently run the sim-

130 ulation at an acceptable computational cost. The period of assimilation for the IS4D-Var cy-

131 cles is four days, which corresponds to the limit of the linearity assumption within the domain

132 [*Matthews et al.*, 2011]. The atmospheric forcing is adjusted every six hours, while the bound-

133 aries are every $12h$. An analysis of these adjustments is performed in Section 6.

134 Eight day forecasts are performed from the end of each cycle using the assimilated state

135 as initial conditions, and the short-range (1-4 days) and mid-range (5-8 days) forecasts are eval-

136 uated for skill.

2.3 Observations

138 Observational data used within this study include satellite measurements of the ocean

139 surface of temperature, height, and salinity, *in situ* depth profiles of temperature and salinity,

140 and surface HFR velocities from High Frequency Radar. Observations close to the boundary

141 and in shallow water are neglected.

2.3.1 Satellite Derived Measurements

143 Sea Surface Temperature (SST) observations are available from two sources at differ-

144 ent time periods: initially we used the Global Ocean Data Assimilation Experiment High Res-

145 olution Sea Surface Temperature (GHRSST) Level 4 OSTIA Global Foundation Sea Surface

146 Temperature Analysis [*PO.DAAC*, 2005], referred to as OSTIA for this work. The data is dis-

147 tributed by the Physical Oceanography Distributed Active Archive Center (PO.DAAC), using

148 optimal interpolation to combine data from the Advanced Very High Resolution Radiometer

149 (AVHRR),the Advanced Along Track Scanning Radiometer (AATSR), the Spinning Enhanced

150 Visible and Infrared Imager (SEVIRI), the Advanced Microwave Scanning Radiometer-EOS

Geoscientific
Model Development
Discussions

Open Access

EGU

151 (AMSRE), the Tropical Rainfall Measuring Mission Microwave Imager (TMI), and *in situ* data.

152 This distribution provides a highly smoothed daily gridded global dataset at the surface at a

153 6km spatial resolution, accurate between $0.2 - 0.5\,°C$ in the domain.

154 Beginning in April 2008, we switched to using the GHRSST Level 4 K10_SST Global

155 1 meter Sea Surface Temperature Analysis data set [*PO.DAAC*, 2008], produced by the Naval

156 Oceanographic Office, and is referred to as NAVO for this work. Also distributed by PO.DAAC,

157 this product combines, in a weighted average, data from AVHRR, AMSRE and the Geosta-

158 tionary Operational Environmental Satellite (GOES) Imager. This distribution provides a daily

159 gridded global dataset at 1 meter depth at a 10km spatial resolution, accurate to $0.4\,°C$ in the

160 domain.

161 Sea Surface Height (SSH) observations are derived using sea level anomaly data from

162 the Archiving, Validation and Interpretation of Satellite Oceanographic data (AVISO) delayed

163 time along track information. The data comes from multiple altimeter satellites measuring the

164 anomaly with respect to a twenty-year mean SSH, homogenized against one of the missions

165 to ensure consistency. Each track has approximately 7km spatial resolution and will usually

166 make multiple passes through our domain each day. To convert from sea level anomaly to sea

167 surface height we add the mean SSH field taken from the six-year model run described early,

168 to which we add the barotropic tidal prediction from TPXO. The accuracy of the swaths de-

169 pend on the source satellite and range from $5-7$cm. We use the AVISO product that has been

170 fully filtered and quality controlled until May 2016. At the time of the experiment, the delayed

171 time data were unavailable beyond May 2016, so the near real-time data were used.

172 Sea Surface Salinity (SSS) data are taken from Aquarius missions daily L3 gridded data

173 set [*PO.DAAC*, 2015] distributed by PO.DAAC. The satellite uses a combination of radiome-

174 ters and scatterometers to estimate the surface salinity, mapped to a coarse $1°$ resolution. Er-

175 rors for this product are around 0.2 ppt. Data for this product are available from August 2011

176 until June 2015.

2.3.2 In Situ *Measurements*

178 Depth profiles of temperature and salinity are obtained from threes sources: the Hawaii

179 Ocean Time-Series (HOT) shipboard Conductivity Temperature Depth (CTD) casts, the global

180 network of Argo floats, and autonomous SeaGliders operated by the University of Hawaii.

181 The Hawaii Ocean Time-Series (HOT) project conducts monthly cruises to the deep wa-

182 ter A Long-term Oligotrophic Habitat Assessment (ALOHA) station (located at 23° 45'N, 158°

Geoscientific

Model Development

Discussions

183 00'W) in order to develop a long term data set of physical and biochemical ocean informa-

184 tion CTD stations of temperature and salinity are concentrated in the region around the sta-

185 tion; although some are also established along the ship route.

186 HOT also conducts regular SeaGlider missions departing from station ALOHA. In ad-

187 dition, PacIOOS conducts occasional SeaGlider surveys in areas close to the south coast of

188 Oahu. The buoyancy driven autonomous underwater vehicles take profiles and transects at depth

189 of temperature and salinity.

190 Observations from the global Argo float network are available from the Argo array Net-

191 work [*USGODAE*, 2016]. The free-drifting floats profile temperature and salinity during as-

192 cension and descension every 10 days of depths down to 2000m [*Oka and Ando*, 2004]. Argo

193 measurements tend to occur in the model domain at a rate of about 1-2 profiles per day.

194 Representational errors for HOT CTDs, Argo Floats, and SeaGliders are defined by the

195 variance of observational data from all available sources across the Pacific sorted into depth

196 bins. These profiles resemble a typical temperature/salinity profile, with a peak temperature

197 error of $0.8K$, and peak salinity error of $0.15ppt$ occurring in the mix layer at a depth around

198 100m.

2.3.3 High Frequency Radar Measurements

200 HFR measurements of surface currents are available from PacIOOS at seven sites around

201 the Hawaiian islands: five around the south-west of Oahu and two on the east coast of the Big

202 Island. Data is available from the first site in October, 2010 with the other sites coming on-

203 line at various times, the most recent being October, 2015. The range for the HFRs on Oahu

204 extend approximately 150km from the coast, while the two Big Island sites are focused on cur-

205 rents around Hilo bay and have a shorter range. At the range limits, HFR data is less reliable.

206 Figure 2 shows the percentage availability of data in the region. Data is counted as available

207 if there is more than 80% in a given week.

208 Both spatially and temporally, the resolution for all sites is significantly higher than the

209 model resolution. The HFR data are low-pass filtered at 3 hours to remove the high frequency

210 signals that may not be resolved by the model (atmospheric forcing fields are every 3 hours).

211 We then provide the spatial field of data every 3 hours. The associated error is calculated in-

212 dividually for each spatial point as the accuracy of the measurements is determined by the lev-

213 els of interference, which increases with range. For each observation point we calculate the
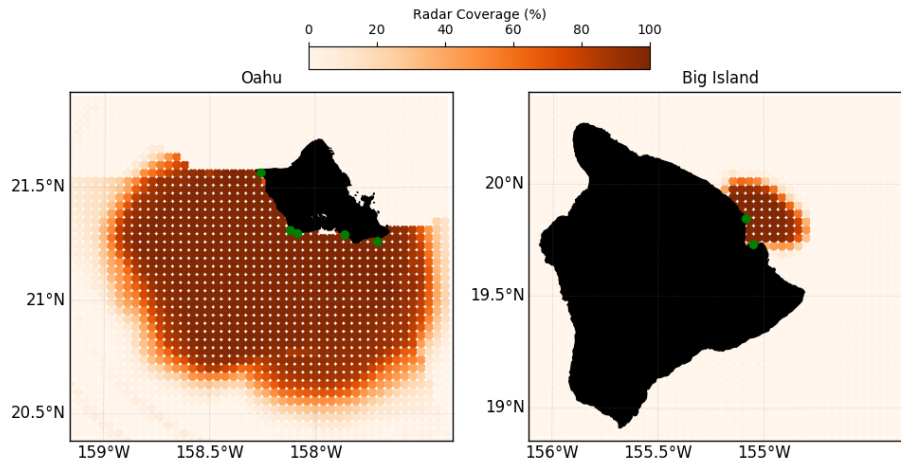
Figure 2. Composite image of percentage coverage for all radar sites (situated at green dots) when all are operational. Where two sites overlap the greater value is taken to indicate the level of coverage at each point.

power spectral density and calculate the noise as per *Zanife et al.* [2003], with a minimum of 7cm/s. At the extreme, errors may reach 17cm/s.

The number of observations for each four day cycle from all sources are shown in Figure 3. Sea surface temperature measurements from both OSTIA and NAVO are consistently the most available observation source, and by the end of the time period HFR is supplying a similar quantity. *In situ* measurements, which include both temperature and salinity for each of the instruments, provide a smaller amount of data by an order of magnitude.

3 Assimilation Statistics

In this section we examine the state estimate to quantify the performance during our time period.
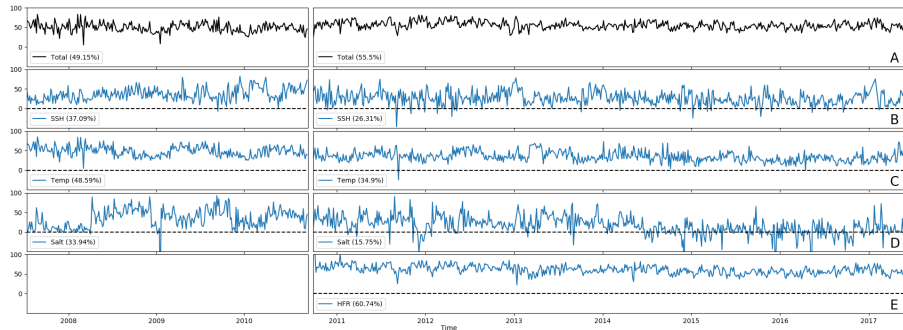
3.1 Cost Function Reduction

I4D-Var minimizes the residuals between the model and observations over each 4-day cycle. We calculate the percentage reduction between the initial and final cost function for each cycle to assess how the assimilation performs over time. Additionally, we can breakdown the cost function further to examine in detail which observation types are reduced the most, however is should be noted that this breakdown does not distinguish between observation sources.
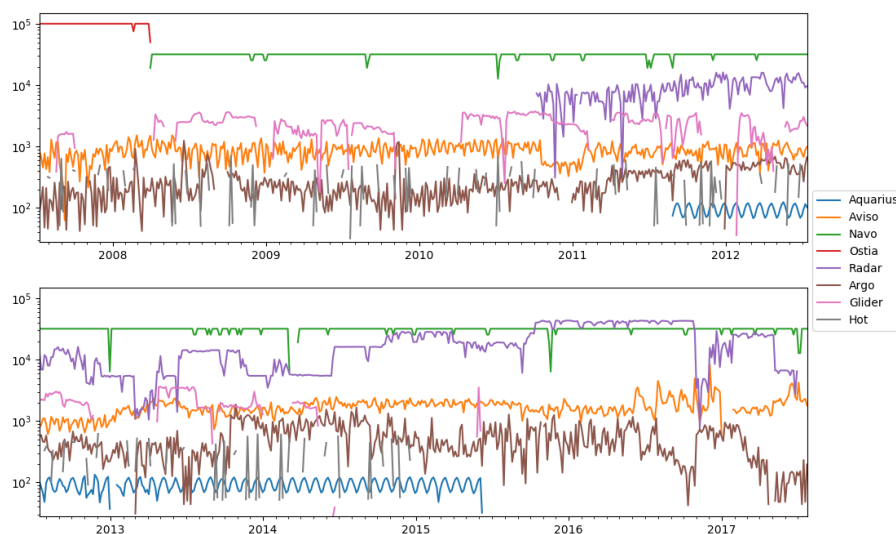
Figure 3. Number of observations used within data assimilation run. Note that there tend to be orders of magnitude more satellite or remotely-sensed observations than *in situ*.

Figure 4 shows a time series of this percentage reduction in the cost function for each of the variables we observe: sea surface height, temperature, salinity and HFR, in addition to the total reduction. A value of 0 means the final cost function is the same as the initial and no reduction has occurred. The plot is split into two distinct time periods, before and after the HFR observations are introduced in order to assess changes in the relative contributions of each variable to the overall reduction.

The total cost function of all data (Figure 4A) is on average halved for each cycle, with an improvement from 49% of the original value to 55% when HFR observations are available. Looking at the breakdown in Figure 4B-E we see that the final cost function associated with the other observed variables: sea surface height, temperature, and salinity, is reduced by a smaller percentage than before HFR was included, which might be expected given the quantity of HFR observations. Salt measurements tend to contribute the least improvement (34% pre-HFR, 16% post-HFR) as they are the least numerous. In some cases, the percent reduction of the cost function for salt is negative, particularly after the Aquarius satellite mission has ended and there were fewer *in situ* observations used. The cost function associated with HFR measurements is reduced by 60% of the initial value, meaning the model is significantly closer to the HFR observations after the assimilation.

Geoscientific
Model Development
Discussions

Figure 4. Time-Series of percentage reduction in the I4D-Var cost function; Left column are pre-HFR observations, right post-HFR, with the mean value given in parentheses. Dashed lines mark the limit of 0, below which there is no reduction in the cost function for that variable. A) Total cost function reduction for all observations; B) Sea surface height observations, C) Temperature observations; D) Salinity observations; E) HFR observations.
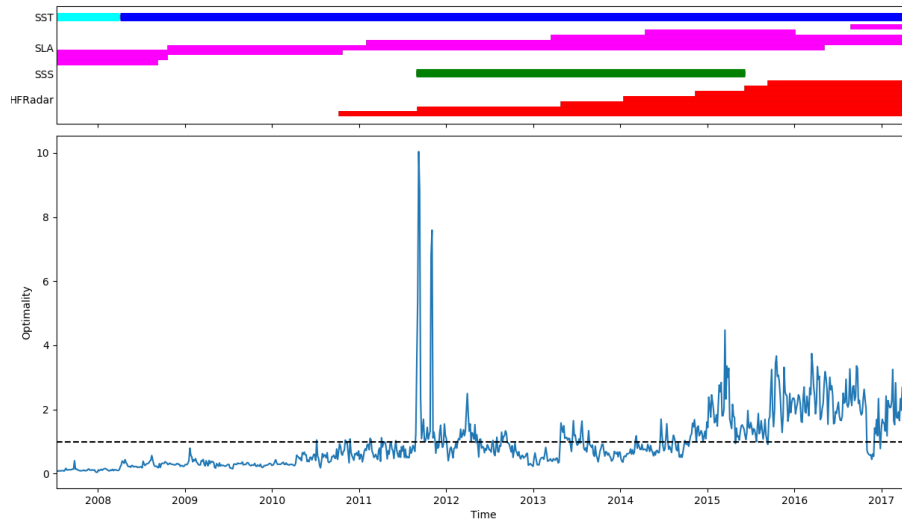
3.2 Optimality

Another measure of the performance is the theoretical minimum value of the cost function [*Bennett*, 2002; *Powell et al.*, 2008], which states that:

$$\frac{2J_{min}}{N_{obs}} = 1, \tag{1}$$

where N_{obs} is the number of observations. Assuming the conjugate gradient algorithm converges, Equation (1) provides a simple representation of how consistently the errors (**P** and **R**) are specified, since the error covariances define the cost function. Figure 5 shows a time-series of the calculated optimality value for the model run, in addition to a timeline of the availability of certain observations for reference. Over the full time period the mean optimality is 0.95; however there are large differences over the course of the time period. In the pre-HFR period the optimality is low, suggesting that the error bounds on observations are too wide. Since SST is the dominant source of observations before HFR, the prescribed errors associated with SST may be too large.

Post-HFR the optimality value increases, suggesting the errors in this period are underestimated. A large optimality value arises when the cost function is large, i.e. large differences between the model and observations. There were two anomalous cycles in 2011, the first of which coincides with the introduction of a second radar site. From 2012 onwards the optimality value is generally good, if highly variable. The trend of increase in optimality given the

Geoscientific
Model Development
Discussions

Figure 5. Top - Gantt chart of remotely sensed observations used in the study. Bottom - Optimality of 4D-Var Assimilation Experiment with the dashed line representing the theoretical minimum.

available observations points to an underestimation of HFR errors, or at the least a persistent difference between the model and HFR observations.

3.3 Error Consistency

A more detailed examination of the consistency of the errors follows the diagnostics described in [*Moore et al.*, 2011b; *Matthews et al.*, 2012]. If the variances in $\mathbf{P}$ and $\mathbf{R}$ are correctly specified *a priori*, they will be consistent with the *a posteriori* estimated via:

$$\widetilde{(\sigma_i^b)}^2 = \frac{1}{p_i} \sum_{j=1}^{p_i} (\mathcal{H}_j(\mathbf{x}^a) - \mathcal{H}_j(\mathbf{x}^b))(y_j - \mathcal{H}_j(\mathbf{x}^b)) \quad \text{and} \tag{2}$$

$$\widetilde{(\sigma_i^o)}^2 = \frac{1}{p_i} \sum_{j=1}^{p_i} (y_j - \mathcal{H}_j(\mathbf{x}^a))(y_j - \mathcal{H}_j(\mathbf{x}^b)), \tag{3}$$

where i refers to the observation type, p_i is the number of observations of that time, y_j is the value of observation j of type i, and $\mathcal{H}_j$ is the mapping of the model background ($\mathbf{x}^b$) or analysis ($\mathbf{x}^a$) to observation j.

Figure 6 shows both the *a priori* and *a posteriori* errors for the remotely sensed data. The observation *a priori* values are calculated as the mean error of the observations in each cycle, while the background *a priori* values are defined as the variability of a free running nonlinear model. If the *a posteriori* errors are typically larger then the *a priori*, it implies the ini-
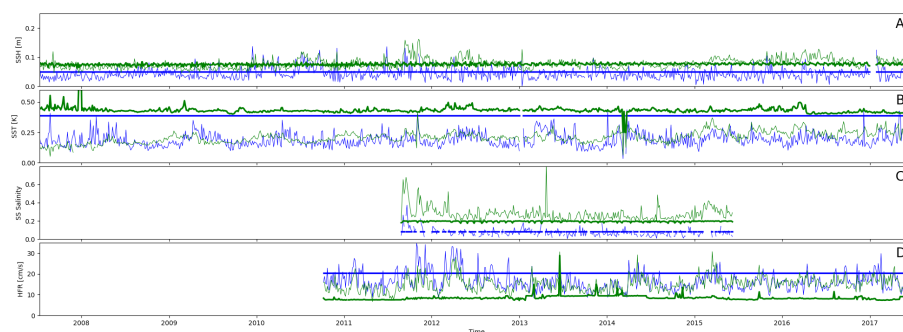
Figure 6. Time series of spatially averaged background (blue) and observation (green) errors, with thick lines showing *a priori* values and thin lines the posterior calculated using Equations (2) and (3). A) Sea Surface Height; B) Sea Surface Temperature; C) Sea Surface Salinity and D) HFR.

tial errors in $\mathbf{P}$ and $\mathbf{R}$ are underestimated. Conversely, if they are smaller the initial errors are likely overestimated.

Figure 6A shows that sea surface height errors are consistent, while sea surface temperature, Figure 6B suggests the *a priori* errors are overestimated. The *a priori* observation errors for NAVO SST observations are defined with a minimum error of 0.4K, but the *a posteriori* are more typically around 0.25K. The *a priori* background errors also also appear overestimated.

Sea surface salinity observation errors (fig. 6C) are slightly underestimated but generally consistent, as are the background errors. The HFR observation errors (fig. 6D) also appear to be underestimated, with most *a priori* errors close to the minimum value of 7cm/s. The *a posteriori* errors suggest a typical value of around $12 - 15$cm/s would be more appropriate. The *a priori* background errors are reasonably consistent with the *a posteriori*, if anything they are slightly overestimated.

This error consistency analysis supports the conclusions in Section 3.2 that the SST observation errors are overestimated and HFR values are underestimated. It is worth noting that these diagnostics are only estimates used to characterize the errors and are not the true posterior error.

Geoscientific
Model Development
Discussions

Open Access

EGU

4 Comparison with Observations

Because I4D-Var relies on the model physics to represent observations through time, it should provide better forecasts. Other methods that perturb the state at single times may better reduce the time-fixed cost function, but can add non-physical structures that generate noisy forecasts.

In this section we examine the state estimate solution by comparing the model to observations. For reference, the observations are also compared against a free-running forecast starting from the same time as each state-estimate cycle. All boundary, atmospheric and tidal forcings are initially the same between runs; however, the boundary and atmospheric forcing are altered as part of the state estimate solution. For comparing fields we use the Root Mean Squared Anomaly (RMSA) and the Anomaly Correlation Coefficient (ACC), defined as:

$$\text{RMSA}(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left((x_i - \bar{x}) - (y_i - \bar{y}) \right)^2} \qquad (4)$$

and

$$\text{ACC}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \qquad (5)$$

where N is the number of observations and x are the model values at the same location and time as the observations y. The RMSA provides a measure of the residual between the model and observations, while the ACC determines the strength of the relationship between the two. We can calculate values for a single spatial point throughout time, or for all spatial points at a single time; however, we require there must be at least 20 observation values available to get a representative statistic. The gridded satellite products are ideally suited to this analysis, while the depth profiles from *in situ* measurements are binned into 50m depth layers to ensure a minimum number of values.
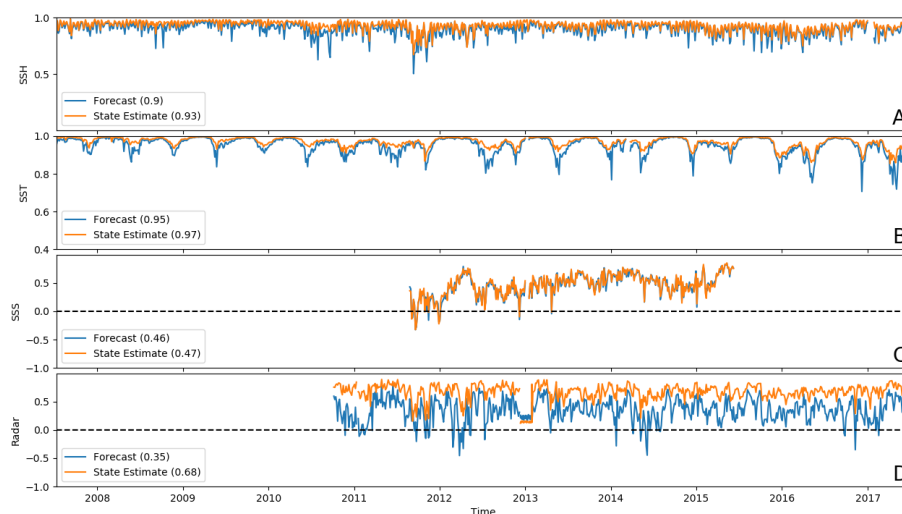
4.1 Remotely Sensed Observations

Figure 7 shows the RMSA between the observations and the models for each source of remotely observed data. The state-estimate solution reduces the RMSA compared with the free-running forecast by 1.58cm (17%), 0.07K (24%), 0.01PSU (3%) and 8.39cm/s (37%) for sea surface height, sea surface temperature, sea surface salinity and HFR respectively. In Figure 7A the RMSA of the state-estimate solution is close to the typical observational error of 7cm, while in Figure 7B we see the RMSA is comfortably less than the 0.4K representative error. Sea surface salinity is only marginally improved by the state-estimate solution, but is slightly over the prescribed observational error of 0.2PSU. The RMSA of the currents associated with HFR
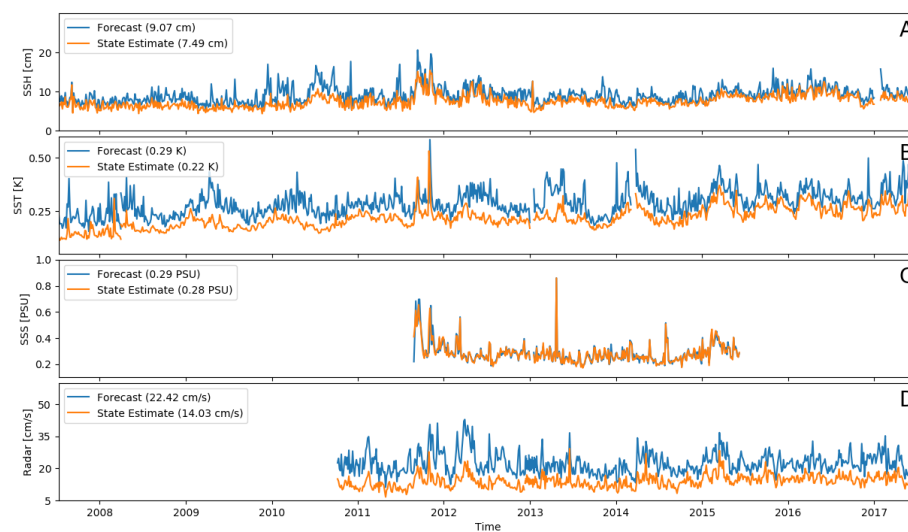
Figure 7. Time series of root mean squared anomalies (RMSA) between remotely sensed observations and two model realizations; the state estimate and a free running forecast. A) Sea Surface Height; B) Sea Surface Temperature; c) Sea Surface Salinity and D) HFRs

observations, shown in Figure 7D, is improved greatly by the state-estimation; however, the mean value of 14cm is around double the typical error prescribed *a priori* of 7cm. As shown in the previous sections, this error was underestimated.

The ACC is also improved by the state-estimate for all variables, as shown in Figure 8. For sea surface height, temperature and salinity the improvement is small due to a significant agreement in the forecast with gains of 0.03, 0.02, and 0.01 respectively. The improvement in HFR is much more significant, with an average correlation improvement from 0.35 to 0.68. As shown in Figure 8D the free-running forecast model can diverge from the observations enough to become negatively correlated over a cycle, while the state-estimate solution is consistently positively correlated. Figure 9 shows the spatial RMSA between the forecast and analyses model solutions and the observations for both sources of sea surface temperature observations: OS-TIA and NAVO. In both cases there is a clear reduction in the RMSA, with the largest source of error in the areas leeward of the islands, most notably the island of Hawai'i. This is due to higher heat flux variability from a reduction in cloud cover [*Yang et al.*, 2008b; *Matthews et al.*, 2012]. Even in this peak area, the state-estimate solution is around the observational error of representativeness of 0.4K, meaning the model is performing well with regards to SST.

Figure 8. Time series of anomaly correlation coefficients (ACC) between remotely sensed observations and two model realizations; the state estimate and a free running forecast. A) Sea Surface Height; B) Sea Surface Temperature; c) Sea Surface Salinity and D) HFRs

Both RMSA and ACC between the experiments and HFR observations are shown in Figure 10 for the island of O'ahu. The RMSA of the free-running forecast is reasonably uniform across the region covered by the HFR, around $20-25$cm/s with some varying values around the extent of the radar coverage. The inclusion of HFR observations in the state-estimate solution leads to significantly reduced values of $12-15$cm/s, a reduction of almost half. The ACC is also significantly improved from a weak correlation to a consistently strong positive one.

As discussed in *Souza et al.* [2015], there are several reasons the model can differ from surface current observations: the discretization of the model, imperfect stratification, differing barotropic-to-baroclinic tide conversion at Kaena ridge, or mixing parameters that do not capture the real baroclinic mixing. This may lead to a different location of the currents in the model from those observed by the HFR; however, the model does a good job reducing these errors [*Janeković and Powell*, 2012]. The HFRs located on the island of Hawai'i have a smaller coverage region, but the level of improvement from the forecast to the state-estimate solution is consistent with the O'ahu results shown here.

.
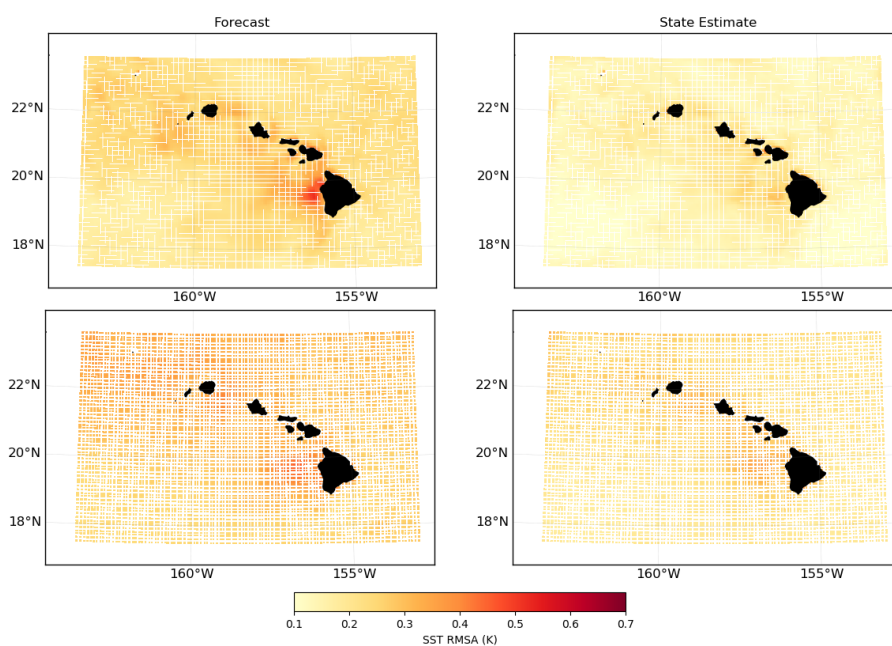
Figure 9. Spatial maps of RMSA for SST observation sources. Top - OSTIA data (2007-2008); Bottom - NAVO data (2008-2017). Typical error of representativeness is around 0.4K.

4.2 Subsurface Observations

The *in situ* observation sources: Argo floats, Seagliders and HOT CTDs also show an improvement in the state estimate over the forecast. The subsurface temperature RMSA values are reduced by an average of 0.03K (5%) and salinity by 0.01PSU (9%). The average values are within the bounds of the maximum representative errors for both variables, 0.8K and 0.15PSU; however, there are several occasions when the RMSA value for a cycle exceeds that limit. In those cases the state-estimate is reducing the cost, but there are too few observations to weight the model towards.

Figure 11 shows the RMSA and ACC profiles for temperature and salinity respectively for each source of subsurface observation. For all three sources, the greatest RMSA between the models and observations is along the thermocline where minor differences in thermocline depth leads to temperature differences. The state-estimate improves the RMSA in this region by $10-15\%$. The thermocline location is also the source of lowest correlation between the
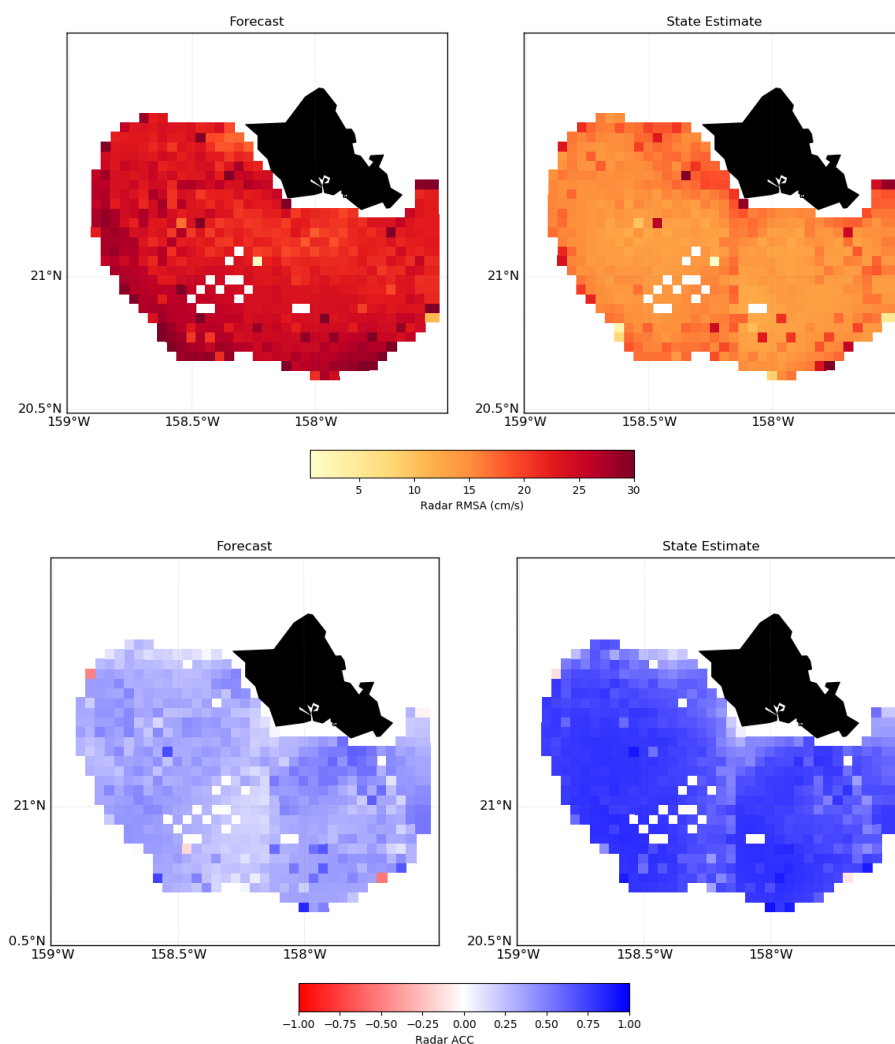
Figure 10. Spatial maps of HFR statistics for south O'ahu. Top - RMSA; Bottom - ACC.

observations and the model, which is also improved by the state-estimate by $\sim 5\%$. There

is a high RMSA for SeaGliders at the base of their profiles (close to 1000m), in this instance

the state-estimate performs worse than the forecast. Many of the Glider missions used oper-

ate in the shallow waters off the south coast of O'ahu and there are very few observations at

this depth compared with higher up the profile.

For subsurface salinity (fig. 11), the improvements made by the state-estimate solution

occur almost exclusively about 500m for Argo floats and HOT CTDs. As with temperature
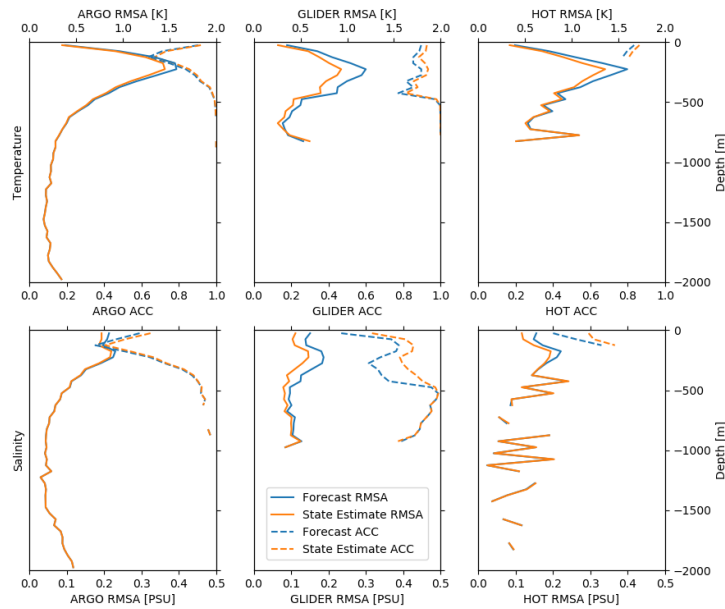
Figure 11. RMSA (solid) and ACC (dashed) profiles, binned by $50m$ intervals, of subsurface temperature (top) and salinity (bottom) for Argo floats, SeaGliders and HOT CTDs.

the largest improvement is at the top of the thermocline. There are some low ACC values lower down in the profile between both models and the observations, but both the forecast and state-estimate perform equally at this depth. SeaGliders produce the biggest improvement in sub-surface salinity model performance, with the state-estimate solution up to 20% better than the forecast for both RMSA and ACC. The peak improvement is at the top of the thermocline, but there are improvements throughout the profile.

5 Forecast Skill

In this section we quantify the model skill by using a skill score (SS), evaluated as the improvement against a reference field [*Murphy*, 1988]. For the reference, we use the persis-tence assumption and take the spatial location of each observation from the model field at the time of initialization for each cycle. The SS for the state estimate and forecast are defined as:

$$\text{SS}_a = 1 - \frac{\text{RMSA}(\mathbf{x}^a, \mathbf{y})}{\text{RMSA}(\mathbf{x}^0, \mathbf{y})}, \tag{6}$$

$$\text{SS}_f = 1 - \frac{\text{RMSA}(\mathbf{x}^f, \mathbf{y})}{\text{RMSA}(\mathbf{x}^0, \mathbf{y})}, \tag{7}$$

where the superscripts a, f, and 0 refer to the state estimate, model forecast and persistence, respectively. Under this measure, a SS of 1 represents a perfect fit between the model and observations, while a value of zero indicates where the model and persistence values perform exactly the same. If the model is better than persistence, then the skill score will lie in the range $0 < \text{SS} < 1$ and the degree of improvement over persistence is determined by how close to 1 the score is. Conversely, a negative SS means the model is further from the observations than persistence.
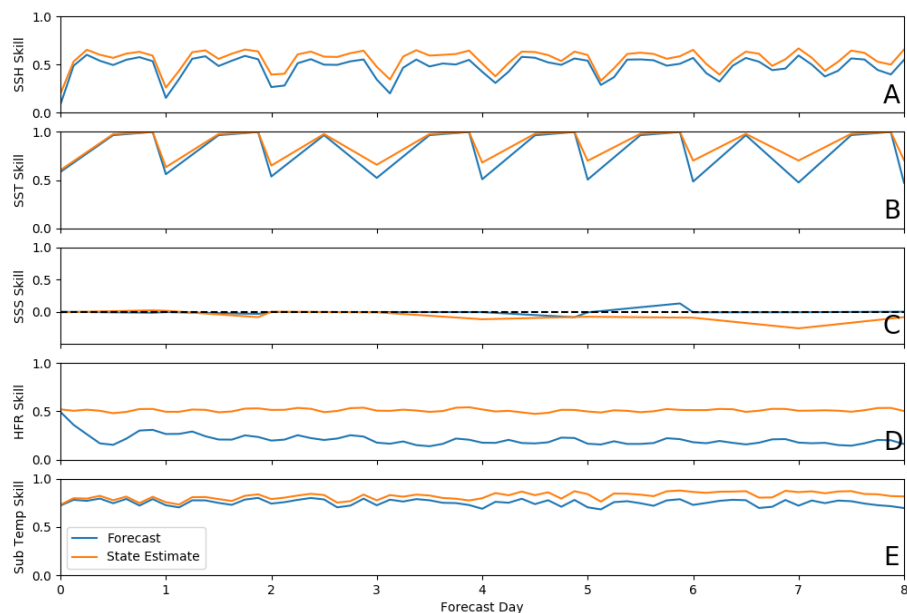
For this verification we wish to examine the effect of forecast length on the skill. Starting with the same initial conditions as each state estimate cycle we produce an eight day forecast, the length of two state estimate cycles. The RMSA is calculated every 3 hours for each 8-day forecast, the corresponding state-estimate cycles, and the persistence field from the start of the forecast.

Figure 12 shows the mean SS over all cycles for remotely sensed observations. For SSH, SST and HFR, the skill for both the state-estimation and free-running forecast is positive throughout, indicating that both models are successful over persistence in representing those variables. SSS however is close to zero and slightly negative meaning the models provide no better information than persistence. SST values are consistently the highest, with a reduction in skill versus persistence for both models once per day. This is expected as the diurnal cycle means the ocean temperature will be close to persistence once per day. For HFR the state-estimate skill has a consistent value of 0.5 regardless of forecast day, while the forecast is closer to 0.2.

6 Analysis of Increments

During each I4D-Var 4-day window, the initial model field, as well as time-varying boundary and surface forcings are adjusted to minimize the residuals. The initial condition increments form a single record for each cycle, while the boundary and surface forcings are perturbed every time they are applied to the model. The perturbations applied to the boundary exhibit only a minor influence on the model (not shown), likely due to the sponge layer dampening the signal. We focus our analysis on the increments of the initial conditions and the surface forcing.

Over the 10 year reanalysis, there are 917 analysis cycles with sixteen surface forcing adjustments (four per day) per cycle. We will examine the modes of variability of the increments via Empirical Orthogonal Functions (EOFs) [*Hannachi*, 2004].

Figure 12. Mean skill metric for remotely sensed observations as a function of forecast length. A) Sea
Surface Height; B) Sea Surface Temperature; c) Sea Surface Salinity and D) HFRs

6.1 Initial Condition Perturbations

For each cycle, the initial perturbation of five main model variables are examined: sea
surface height, temperature, salinity, east-west velocity and north-south velocity. With the ex-
ception of sea surface height, each variable is averaged over the upper 100m to cover the mixed
layer. The increments for salinity and sea surface height as a percentage of the background
values are insignificant ($< 1\%$), while temperature increments ($2-10\%$) and the two veloc-
ity fields ($10-20\%$) are significant enough to analyze.

Figure 13 shows the first three EOF modes for temperature and velocity increments to
the initial conditions. For temperature, the first mode accounts for 32% of the variance and
is positive across the entire domain with a peak in the region leeward of the islands, an area
known to have a strong temperature gradient [*Xie et al.*, 2001]. The second and third modes,
($\sim 11-13\%$ of variability) are also strongest in this region, albeit much weaker.

The first three modes of the velocity increments are closer in explained variance than
temperature, with the first accounting for $\sim 25\%$ and the second and third $\sim 15\%$. The EOFs
exhibit a stronger signal further south than temperature, where the North Equatorial Current
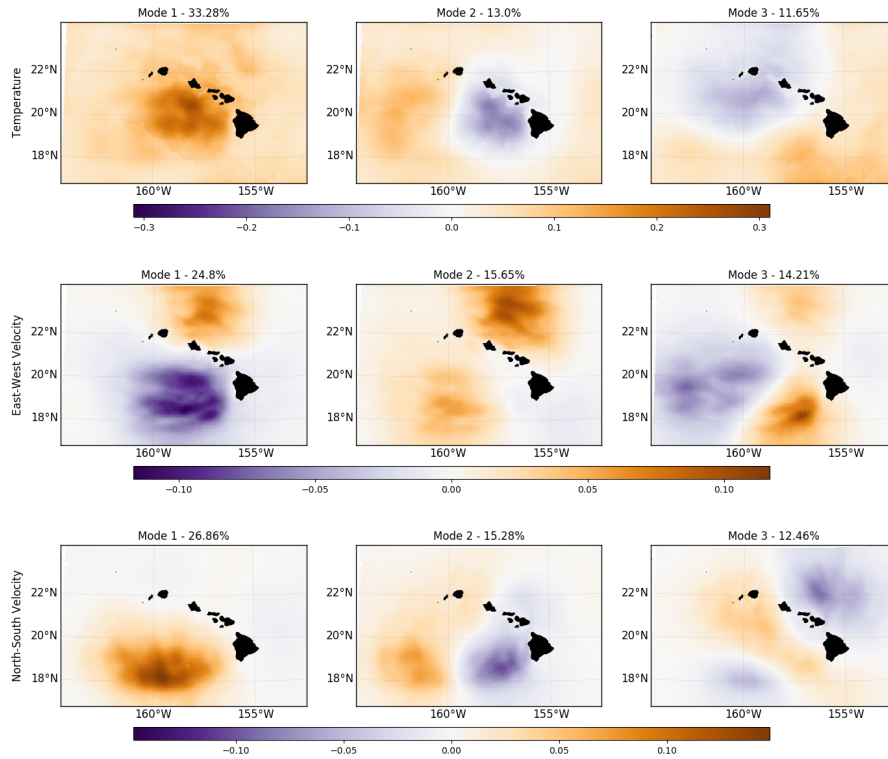
Figure 13. EOF analysis of initial condition increments for upper 100m: top - temperature, middle - east-west velocity and bottom - north-south velocity.

(NEC) flows past the islands and passes the Hawaiian Lee Counter Current (HLCC). Look-ing at the east-west and north-south velocities in tandem, we can see that the dominant mode is increasing the velocity to the north-west in this region. The secondary mode is weak in that area for the east-west direction, with a split signal in the north south direction, while the third mode shows the opposite. These modes are adjusting the state estimate for anti-cyclonic eddy activity in lee of the Big Island.
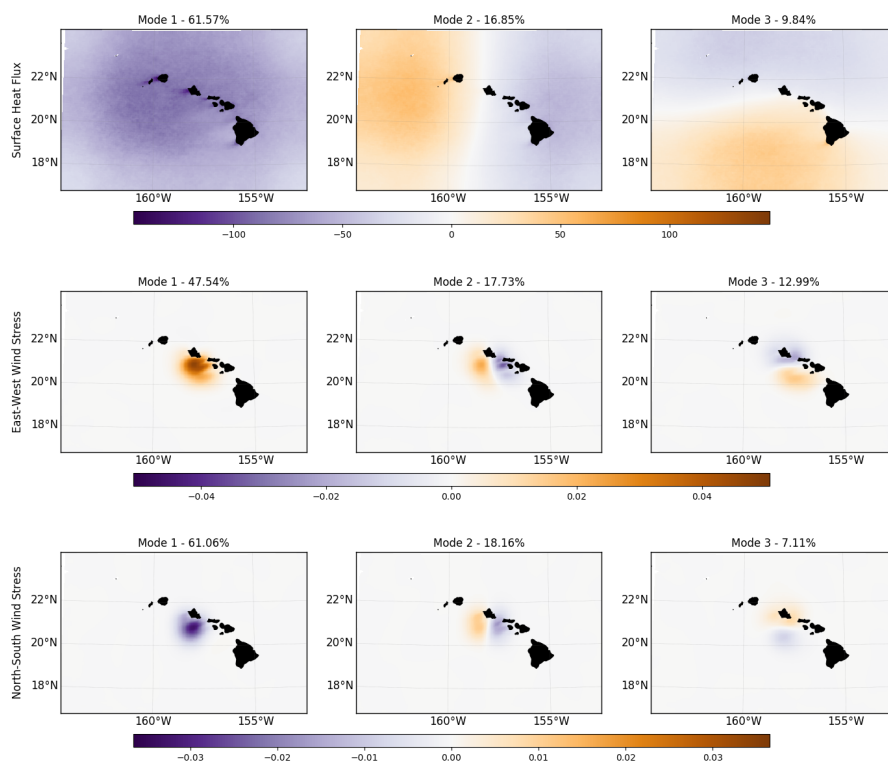
6.2 Surface Forcing Perturbations

The surface forcing increments are calculated every 6 hours during the assimilation and are linearly interpolated between this period. The time of day potentially impacts forcing variables, particularly surface heat flux, so we calculate EOFs on the increments for each of the four distinct times of day they occur (00, 06, 12, 18 UTC). Due to the size of the model grid, the number of records and the computational resources available the EOF calculation is limited to a 4-year period, approximately 1500 records. The results presented are the EOFs for the period 2007-2011; however other 4-year periods were also examined with no significant differences in the structure of the modes or their percentage variance explained. The time of day does impact the percentage variance explained by each mode, most notably for surface heat flux where the effect of diurnal solar heating occurs. The composition of the modes themselves (i.e. where the peak/trough variances are located), have no significant difference for each time of day, so we present the average modes.

The four key surface forcing terms are: surface heat flux, surface salinity flux, east-west wind stress, and north-south wind stress. Of these, increments in surface salinity flux are small compared to their background value, while increments in surface heat flux ($10-15\%$ of background) and the wind stresses ($15-20\%$) are significant. The first mode dominates the majority of the variability, particularly for surface heat flux where 60% is explained by this mode alone. For all three variables, the majority of the variability can be explained using the first 5 modes.

The first three modes for the forcing terms are shown in Figure 14. For surface heat flux, we observe that the strongest mode, representing nearly 60% of the variability, is consistently negative over the region, essentially just accounting for the bias with the model. The strongest low in the region is leeward of the islands, similar to the first mode for the initial ocean temperature field and indicative that SST is being controlled by the surface heat flux. The second and third modes (16% and 10% of variance), split the domain vertically and horizontally, with a much weaker signal than the dominant mode. These two modes allow for large scale atmospheric gradients to contribute towards the surface heat forcing.

The EOFs of surface wind stress increments are confined to the shallow region close to the south coast of O'ahu, where the surface winds are likely to have the most impact on HFR. The strong singular signal from the dominant modes show a south-east high to adjust the winds; whereas, the secondary modes shows a diverging north-west to south-east split in the region account for more minor lee affects. The same signal does not appear in the initial condition

Figure 14. EOF analysis of atmospheric forcing perturbations; a) surface heat flux, b) east-west wind stress and c) north-south wind stress.

perturbations for velocity, likely because the wind stress has a significant effect on the surface flow.

7 Summary

We have presented a 10-year reanalysis of the PacIOOS Hawaiian Island Ocean Forecast System and assessed the performance of the state-estimate solution and free-running forecasts. Using an updated model and data assimilation scheme, we show that the model represents the observational data well over the time period. The largest improvement in the state-estimate solution occurs when minimizing the residuals to HFR data, with sea surface temperature also accounting for a significant improvement. On average, the assimilation achieves the near-optimal solution; however, the variability is heavily influenced by the HFR observations. The analysis suggests the observational errors associated with HFR are too low and results could be improved by redefining these errors. This is supported by the increase in variability and upward trend of optimality towards the end of the time period where HFR observations are most numerous.

The increments made by the reanalysis have revealed that sea surface height and salinity initial conditions aren't significantly adjusted by the I4D-Var procedure; whereas temperature and velocity account for a significant change from the forecast field. For the atmospheric forcing, surface salinity is insignificant, but the surface heat flux and wind stresses adjust the forcings by up to 20%. This corresponds to cost function statistics that show HFR and temperature as the two dominant observation sources. For each of these increments we performed an EOF analysis, where the first few modes explain most of the variability. Typically the first mode is dominant; In the atmospheric forcing increments the first mode is much more significant than the others and tends to have a uniform direction indicating a bias adjustment. For surface heat flux this applies across the whole domain while wind stress increments are concentrated in the region south of O'ahu. The wind stress heavily influences the surface currents and are mostly due to HFR. The secondary modes tended to show more oscillatory behavior to allow for atmospheric gradients and more minor effects.

The reanalysis has provided the testing for improvements to the PacIOOS operational forecast system. The data is being used to update the back catalog available publicly at www.pacioos.org and will influence the future results from daily forecasts. Analysis of the I4D-Var increments has provided a greater understanding of the variability in the region and will provide the basis for a move towards ensemble forecasting in the region.

8 Code and Data Availability

ROMS is an open source software package distributed freely from http://www.myroms.org.

Geoscientific
Model Development
Discussions

Open Access

EGU

Model initial conditions and boundary forcing comes from the HYbrid Coordinate Ocean
Model (hycom.org). Atmospheric surface forcing and HFRadar observations are distributed
through the PacIOOS data portal (pacioos.hawaii.edu).

Satellite measurements come from two sources; sea surface temperature and salinity are
provided by the Physical Oceanography Distributed Active Archive Centre (podaac.jpl.nasa.gov),
and surface height anomalies are provided by the Copernicus Marine Environment Monitor-
ing Service (marine.copernicus.eu).

In Situ measurements used are available from 3 sources; Argo measurements through
Global Ocean Data Assimilation Experiment (usgodae.org), SeaGliders through the School of
Ocean and Earth Science and Technology at the University of Hawai'i at Manoa (hahana.soest.hawaii.edu/seagliders),
and CTDs through the Hawaii Ocean Time-Series project (hahana.soest.hawaii.edu/hot).

Reanalysis output is produced as 3-hourly snapshots of the 3D fields temperature, salin-
ity and velocities, as well as the 2D sea surface height field for the full time period. This data
is archived through PacIOOS and can be made available for research purposes.

Acknowledgements

References

Bennett, A. (2002), *Inverse Modeling of the Ocean and Atmosphere*, Cambridge University
    Press, doi:10.1017/CBO9780511535895.

Broquet, G., C. Edwards, A. Moore, B. Powell, M. Veneziani, and J. Doyle (2009), Ap-
    plication of 4d-variational data assimilation to the California current system, *Dynamics
    of Atmospheres and Oceans*, *48*(1–3), 69 – 92, doi:10.1016/j.dynatmoce.2009.03.001,
    modeling and Data Assimilation in Support of Coastal Ocean Observing Systems.

Chapman, D. C. (1985), Numerical treatment of cross-shelf open boundaries in a
    barotropic coastal ocean model, *Journal of Physical Oceanography*, *15*(8), 1060–1075,
    doi:10.1175/1520-0485(1985)015⟨1060:NTOCSO⟩2.0.CO;2.

Chassignet, E. P., H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J.
    Wallcraft, R. Baraille, and R. Bleck (2007), The HYCOM (HYbrid Coordinate Ocean

Geoscientific
Model Development
Discussions

550   Model) data assimilative system, *Journal of Marine Systems*, *65*(1), 60 – 83, doi:

551   10.1016/j.jmarsys.2005.09.016.

552   Chavanne, C., P. Flament, R. Lumpkin, B. Dousset, and A. Bentamy (2002), Scatterometer

553   observations of wind variations induced by oceanic islands: Implications for wind-

554   driven ocean circulation, *Canadian Journal of Remote Sensing*, *28*(3), 466–474, doi:

555   10.5589/m02-047.

556   Courtier, P., J.-N. Thépaut, and A. Hollingsworth (1994), A strategy for operational im-

557   plementation of 4d-var, using an incremental approach, *Quarterly Journal of the Royal*

558   *Meteorological Society*, *120*(519), 1367–1387, doi:10.1002/qj.49712051912.

559   Egbert, G. D., A. F. Bennett, and M. G. G. Foreman (1994), Topex/poseidon tides esti-

560   mated using a global inverse model, *Journal of Geophysical Research: Oceans*, *99*(C12),

561   24,821–24,852, doi:10.1029/94JC01894.

562   Flather, R. (1976), A tidal model of the northwest european continental shelf, *Mem. Soc.*

563   *R. Sci. Liege*, *10*(6), 141–164.

564   Hannachi, A. (2004), A primer for EOF analysis of climate data, *Tech. rep.*, Department of

565   Meteorology, University of Reading.

566   HMRG (2017), Hawaii Mapping Research Group, SOEST. `http://www.soest.`

567   `hawaii.edu/HMRG/multibeam/index.php`, Online, Last Checked 04/13/2018.

568   Janeković, I., and B. S. Powell (2012), Analysis of imposing tidal dynamics to nested

569   numerical models, *Continental Shelf Research*, *34*, 30–40, doi:10.1016/j.csr.2011.11.017.

570   Janeković, I., B. S. Powell, D. Matthews, M. A. McManus, and J. Sevadjian (2013), 4d-

571   var data assimilation in a nested, coastal ocean model: A Hawaiian case study, *Journal*

572   *of Geophysical Research: Oceans*, *118*, 5022–5035, doi:10.1002/jgrc.20389.

573   Kerry, C., B. Powell, M. Roughan, and P. Oke (2016), Development and evaluation of a

574   high-resolution reanalysis of the east australian current region using the regional ocean

575   modelling system (ROMS 3.4) and incremental strong-constraint 4-dimensional varia-

576   tional (is4d-var) data assimilation, *Geoscientific Model Development*, *9*(10), 3779–3801,

577   doi:10.5194/gmd-9-3779-2016.

578   Lumpkin, R., and P. Flament (2013), Extent and energetics of the Hawaiian lee counter-

579   current, *Oceanography*, *26*(1), 58–65, doi:10.5670/oceanog.2013.05.

580   Matthews, D., B. S. Powell, and R. Milliff (2011), Dominant spatial variability scales

581   from observations around the Hawaiian islands, *Deep-Sea Research*, *58*(10), 979–987,

582   doi:10.1016/j.dsr.2011.07.004.

583   Matthews, D., B. S. Powell, and I. Janeković (2012), Analysis of four-dimensional vari-

584   ational state estimation of the Hawaiian waters, *Journal of Geophysical Research:*

585   *Oceans*, *117*, C03,013, doi:10.1029/2011JC007575.

586   Moore, A. M., H. G. Arango, E. D. Lorenzo, B. D. Cornuelle, A. J. Miller, and D. J.

587   Neilson (2004), A comprehensive ocean prediction and analysis system based on the

588   tangent linear and adjoint of a regional ocean model, *Ocean Modelling*, *7*(1–2), 227 –

589   258, doi:10.1016/j.ocemod.2003.11.001.

590   Moore, A. M., H. G. Arango, G. Broquet, B. S. Powell, A. T. Weaver, and J. Zavala-

591   Garay (2011a), The Regional Ocean Modeling System (ROMS) 4-dimensional varia-

592   tional data assimilation systems: Part I – system overview and formulation, *Progress in*

593   *Oceanography*, *91*(1), 34–49, doi:10.1016/j.pocean.2011.05.004.

594   Moore, A. M., H. G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell, D. Foley,

595   J. D. Doyle, D. Costa, and P. Robinson (2011b), The Regional Ocean Modeling System

596   (ROMS) 4-dimensional variational data assimilation systems: Part II – performance and

597   application to the california current system, *Progress in Oceanography*, *91*(1), 50 – 73,

598   doi:10.1016/j.pocean.2011.05.003.

599   Moore, A. M., H. G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell, D. Fo-

600   ley, J. D. Doyle, D. Costa, and P. Robinson (2011c), The Regional Ocean Modeling

601   System (ROMS) 4-dimensional variational data assimilation systems: Part III – obser-

602   vation impact and observation sensitivity in the california current system, *Progress in*

603   *Oceanography*, *91*(1), 74 – 94, doi:10.1016/j.pocean.2011.05.005.

604   Murphy, A. H. (1988), Skill scores based on the mean square error and their relation-

605   ships to the correlation coefficient, *Monthy Weather Review*, *116*(12), 2417–2424, doi:

606   10.1175/1520-0493(1988)116⟨2417:SSBOTM⟩2.0.CO;2.

607   Oka, E., and K. Ando (2004), Stability of temperature and conductivity sensors of

608   argo profiling floats, *Journal of Oceanography*, *60*(2), 253–258, doi:10.1023/B:

609   JOCE.0000038331.10108.79.

610   PacIOOS (2018), Pacific Islands Ocean Observing System. `http://www.pacioos.`

611   `hawaii.edu/`, Online, Last Checked 04/13/2018.

612   Penenko, V. V. (2009), Variational methods of data assimilation and inverse problems for

613   studying the atmosphere, ocean, and environment, *Numerical Analysis and Applications*,

614   *2*(4), 341–351, doi:10.1134/S1995423909040065.

615 PO.DAAC (2005), UK Met Office. 2005. GHRSST level 4 OSTIA global foundation sea
616   surface temperature analysis. ver. 1.0.

617 PO.DAAC (2008), Naval Oceanographic Office. GHRSST level 4 K10 global 1 meter sea
618   surface temperature analysis. ver. 1.0.

619 PO.DAAC (2015), NASA Aquarius project. Aquarius official release level 3 sea surface
620   salinity standard mapped image daily data v4.0.

621 Powell, B., H. Arango, A. Moore, E. D. Lorenzo, R. Milliff, and D. Foley (2008), 4DVAR
622   data assimilation in the Intra-Americas sea with the Regional Ocean Modeling System
623   (ROMS), *Ocean Modelling*, *25*(3–4), 173 – 188, doi:10.1016/j.ocemod.2008.08.002.

624 Sasaki, H., and P. Klein (2012), Ssh wavenumber spectra in the north pacific from a high-
625   resolution realistic simulation, *Journal of Physical Oceanography*, *42*(7), 1233–1241,
626   doi:10.1175/JPO-D-11-0180.1.

627 Shchepetkin, A. F., and J. C. McWilliams (1998), Quasi-monotone advection schemes
628   based on explicit locally adaptive dissipation, *Monthly Weather Review*, *126*(6), 1541–
629   1580, doi:10.1175/1520-0493(1998)126⟨1541:QMASBO⟩2.0.CO;2.

630 Shchepetkin, A. F., and J. C. McWilliams (2003), A method for computing horizontal
631   pressure-gradient force in an oceanic model with a nonaligned vertical coordinate,
632   *Journal of Geophysical Research: Oceans*, *108*(C3), doi:10.1029/2001JC001047, 3090.

633 Shchepetkin, A. F., and J. C. McWilliams (2005), The Regional Oceanic Modeling System
634   (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model,
635   *Ocean Modelling*, *9*(4), 347–404, doi:10.1016/j.ocemod.2004.08.002.

636 Simmons, A. J., J. M. Wallace, and G. W. Branstator (1983), Barotropic wave propagation
637   and instability, and atmospheric teleconnection patterns, *Journal of the Atmospheric*
638   *Sciences*, *40*(6), 1363–1392, doi:10.1175/1520-0469(1983)040⟨1363:BWPAIA⟩2.0.CO;2.

639 Smith, R. B., and V. Grubišić (1993), Aerial observations of Hawaii's wake, *Journal of*
640   *the Atmospheric Sciences*, *50*(22), 3728–3750, doi:10.1175/1520-0469(1993)050⟨3728:
641   AOOHW⟩2.0.CO;2.

642 Souza, J. M. A. C., B. S. Powell, A. C. Castillo-Trujillo, and P. Flament (2015), The
643   vorticity balance of the ocean surface in Hawaii from a regional reanalysis, *Journal of*
644   *Physical Oceanography*, *45*(2), 424–440, doi:10.1175/JPO-D-14-0074.1.

645 Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft,
646   C. N. Hill, and J. Marshall (2002), Global ocean circulation during 1992–1997, esti-
647   mated from ocean observations and a general circulation model, *Journal of Geophysical*

648   *Research: Oceans*, *107*(C9), 1–1–1–27, doi:10.1029/2001JC000888, 3118.

649   Talagrand, O., and P. Courtier (1987), Variational assimilation of meteorological obser-

650   vations with the adjoint vorticity equation. I: Theory, *Quarterly Journal of the Royal*

651   *Meteorological Society*, *113*(478), 1311–1328, doi:10.1002/qj.49711347812.

652   USGODAE (2016), Argo floats data from global data assembly centre

653   doi:10.17882/42182.

654   Weaver, A. T., J. Vialard, and D. L. T. Anderson (2003), Three- and four-dimensional

655   variational assimilation with a general circulation model of the tropical pacific ocean.

656   part i: Formulation, internal diagnostics, and consistency checks, *Monthly Weather*

657   *Review*, *131*(7), 1360–1378, doi:10.1175/1520-0493(2003)131⟨1360:TAFVAW⟩2.0.CO;2.

658   WRF-ARW (2017), Hawaii Weather Research and Forecasting, SOEST. `http://www.`

659   `soest.hawaii.edu/MET/Faculty/wrf/arw/arw_state_6km.html`, On-

660   line, Last Checked 04/13/2018.

661   Xie, S. P., W. T. Liu, Q. Y. Liu, and M. Nonaka (2001), Far-reaching effects of the

662   Hawaiian islands on the pacific ocean-atmosphere system, *Science*, *292*(5524), 2057–

663   2060, doi:10.1126/science.1059781.

664   Yang, Y., Y.-L. Chen, and F. M. Fujioka (2008a), Effects of trade-wind strength and di-

665   rection on the leeside circulations and rainfall of the island of Hawaii, *Monthly Weather*

666   *Review*, *136*(12), 4799–4818, doi:10.1175/2008MWR2365.1.

667   Yang, Y., S.-P. Xie, and J. Hafner (2008b), Cloud patterns lee of Hawaii island: A synthe-

668   sis of satellite observations and numerical simulation, *Journal of Geophysical Research*,

669   *113*(D15), doi:10.1029/2008JD009889.

670   Zanife, O. Z., P. Vincent, L. Amarouche, J. P. Dumont, P. Thibaut, and S. Labroue (2003),

671   Comparison of the Ku-band range noise level and the relative sea-state bias of the

672   Jason-1, TOPEX, and Poseidon-1 radar altimeters, *Marine Geodesy*, *26*(3-4), 201–238,

673   doi:10.1080/714044519.

674   Zhang, W. G., J. L. Wilkin, and H. G. Arango (2010), Towards an integrated ob-

675   servation and modeling system in the New York bight using variational meth-

676   ods. Part I: 4DVAR data assimilation, *Ocean Modelling*, *35*(3), 119 – 133, doi:

677   10.1016/j.ocemod.2010.08.003.