Results from the SOSSTA project on developing a statistical-dynamical observation operator for SST data assimilation

Sam Pimentel, Wang-Hung Tse
Andrea Storto, Eric Jansen, Isabelle Mirouze
Gerasimos Korres, Demitra Denaxa

TWU, Canada
CMCC, Italy
HCMR, Greece

Plenary Session VIII: Diurnal Variability
20th GHRSSST Science Team Meeting, Frascati, Italy, 6 June 2019
The Science Problem

Image modified from ghrsst.org
Aim

• Develop and test a new technique for producing a dynamically-based highly efficient statistical observation operator for the assimilation of satellite SST observations
Aim

- Develop and test a new technique for producing a dynamically-based highly efficient statistical observation operator for the assimilation of satellite SST observations

Solution Approach

- Use a high-resolution process model of the near surface ocean thermodynamics (GOTM) to generate a training data set
- Perform Canonical Correlation Analysis (CCA) to correlate the profile data with skin & sub-skin SST variations
- The resulting statistical observation operator would include diurnal variability conditioned on categories of atmospheric forcing
This work provides a statistical observation operator $H$, that projects the model state $x$ onto the space of the satellite SST observations $y$. The equation is: $y^o = H(x, f) + \varepsilon^o$. The observation operator, i.e., canonical correlations calculated from GOTM, and atmospheric conditions (wind and insolation) and time of day are used in the model state vector (temperature profile). The SST observations (L2) in skin/subskin SST space contribute to this approach.
Modelling Diurnal Variability in SST

• GOTM, a 1-D process model of near-surface ocean thermodynamics tailored for modelling the SST diurnal cycle

• 122 levels to depth of 75m, higher resolution near surface (e.g. top level is 0.015m, 21 levels in top 1m, 52 levels in top 5m, and 68 levels in top 10m)

• Various solar absorption parameterisations tested including using MODIS chlorophyll data (Ohlmann & Siegel, 2000) and MODIS IOP data (Lee et al., 2005)

• Skin SST is computed dynamically within GOTM using the parameterisation (Fairall et al., 1996a) and takes into account the fraction of SWR absorbed in the cool-skin layer
GOTM Mediterranean Sea Simulations

Forcing:

- 3-hourly atmospheric data from ECMWF ERA-Interim: u10m, v10m, t2m, q2m, airp, swrd, lwrd
- Compute air-sea fluxes using linearly interpolated ERA data and modelled skin SST (Fairall et al., 1996a, 1996b, 2003)
- SWR computed every time step, but match 3-hourly integrated ERA-Interim values

Initialization:

- Daily Temperature and Salinity profiles from MED MFC reananalysis are used to initialize GOTM at sunrise each day
Sample GOTM Output

SST April 2013 at 33.75°N 26.25°E
GOTM Results

\[ T_{\text{subskin}} - T_{\text{skin}} \]

\[ T_{\text{subskin}} - T_{1.5m} \]
1/16-degree GOTM Output: Mediterranean Sea

Diurnal Warming in 2013, Day of Year= 2

Animation showing daily diurnal variability: (max SST – min SST)
Canonical Correlation Analysis (CCA)

- CCA calculates a transformation into pairs of variables that are maximally correlated, but uncorrelated to all previous pairs.
- Matrices $a$ and $b$ can be calculated on a training data set (our GOTM simulations), then the transformation $ab^{-1}$ can be applied to project a temperature profile $T$ onto a satellite observation $S$.

\[
\begin{align*}
\text{Model temperature} & \quad T = (T_1, T_2, \ldots, T_N) \\
\text{Satellite observations} & \quad S = (S_1, S_2) \\
\text{Canonical temperature} & \quad U_1 = (a_{11}T_1, a_{12}T_2, \ldots, a_{1N}T_N) \\
& \quad U_2 = (a_{21}T_1, a_{22}T_2, \ldots, a_{2N}T_N) \\
\text{Canonical observations} & \quad V_1 = (b_{11}S_1, b_{12}S_2) \\
& \quad V_2 = (b_{21}S_1, b_{22}S_2) \\
\end{align*}
\]

$U_i$ maximally correlated with $V_i$

$U_i$, $V_i$ uncorrelated with $U_j$, $V_j$ for $j < i$
**Correlation Plots**

**Measurement vs. model:** measurement is correlated to all levels, BUT all levels are also correlated to each other!

**Canonical variables:** first correlation is as strong as on the left, BUT we found an additional correlation of 60%!
Some Examples

- During low wind and high insolation, the daytime skin and subskin SST can be significantly higher than the highest model level.
- The statistical operator can describe these temperatures reasonably well.

Wind 0-2.5m/s, insolation 305-315W/m², time 5-6h

- Wind 0-2.5m/s, insolation 305-315W/m², time 14-15h
Wind 8-10m/s, insolation 305-315W/m², time 14-15h

- Higher wind speeds cause more mixing in the water column and the diurnal heating effect essentially disappears

Wind 8-10m/s, insolation 0-80W/m², time 14-15h

- During days of high wind and low insolation, the skin temperature is evidently below that of the highest model level
Performance: Skill scores

\[ SS = 1 - \frac{MSE_{CCA\ OO}}{MSE_{reference}} \]

- \( MSE \) = mean square error between subskin SEVIRI L3C obs and model SST
- CCA OO = using the subskin SST produced by the CCA observation operator
- Reference = upper MED MFC model level (1.47m)
Performance: Skill scores

\[ SS = 1 - \frac{MSE_{CCA \text{ OO}}}{MSE_{Bernie}} \]

- **MSE** = mean square error between subskin SEVIRI L3C obs and model SST
- **CCA OO** = using the subskin SST produced by the CCA observation operator
- **Bernie** = using the diurnal SST parameterisation of Bernie et al., 2007
## Offline results – Aegean Sea

**Greek POSEIDON SEEK Filter**

### Monthly Evaluation

<table>
<thead>
<tr>
<th>Month</th>
<th>2014 without operator</th>
<th>2014 with operator</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>0.9412</td>
<td>0.8380</td>
<td>11%</td>
</tr>
<tr>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sept</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**RMSE without OO**

**RMSE with OO**
Online results – Aegean Sea

Forecast RMS error

Analysis RMS error

Skin SST

Sea surface height

Temperature (below skin — 1000m)

Weeks →

Weeks →

Control

DA using OO

Control

DA using OO

Control

DA using OO
Conclusions

- GOTM is used to produce a high-resolution data set of diurnal SSTs in the Mediterranean Sea to compare skin SST, subskin SST, SST at depth, and foundation SST

- This training data set was used to produce canonical correlations, which we found to be a simple and efficient means to provide good estimates of the skin and sub-skin SST. The CCA is used to create a low computational cost observation operator for assimilating satellite SST

- The CCA OO has been tried in the POSIEDON Aegean Sea model, finalized results are forthcoming

- Other potential uses of this method: diurnal skin SST for air-sea flux calculations, comparing different in-situ observations to satellite, diurnal SST in climate models, …