National Research Foundation

SAEON South African Environmental Observation Network

**TECHNICAL BRIEF**

# SOMISANA SA-West V1.0:

# Model Evaluation

**Date**

**PREPARED BY:**

**SOMISANA TEAM**

# Table of Contents

# Executive Summary


The SA-West coastal forecast system, which integrates MERCATOR ocean boundary conditions with either GFS or SAWS atmospheric forcing, was evaluated against satellite altimetry and in situ temperature observations to assess its predictive skill along the South African west coast.

The system successfully reproduces key large-scale circulation features, including the Benguela Jet and St. Helena Bay recirculation, particularly within the first 3 days of the forecast. Beyond this period, forecast accuracy decreases, especially for the 5-day simulations.

Validation against subsurface temperature measurements shows that SA-West generally improves upon MERCATOR in capturing coastal variability and reducing temperature bias. The greatest improvements are observed at Cape Hangklip, where SA-West consistently outperforms MERCATOR. At Cape Point, both systems perform well, with SA-West excelling in the short term (up to 3 days) and MERCATOR maintaining skill over longer lead times. At St. Helena Bay, both models show limited skill, although SA-West better represents variability and bias.

Skill score comparisons confirm that, despite overall low absolute scores due to strict validation against observations, SA-West provides a relative performance advantage in most coastal regions, particularly where finer-scale atmospheric forcing is important.

In summary, SA-West offers enhanced short-term forecasting capability and improved representation of coastal processes, with clear benefits for nearshore applications. Further development should focus on extending forecast reliability beyond 3 days and refining the representation of local dynamics.

# 1. Introduction

The South African west coast is a highly productive marine ecosystem that supports valuable fishing and aquaculture industries. The region forms part of the Benguela Current upwelling system, one of the world's major eastern boundary current systems, characterised by wind-driven upwelling of cold, nutrient-rich waters that sustain high levels of biological productivity. This dynamic environment is further shaped by strong seasonal winds, coastal jets, and mesoscale features such as filaments and eddies, which introduce pronounced spatial and temporal variability in circulation and temperature structure.

Coastal ocean forecasting along the South African west coast plays a critical role in supporting environmental management and operational decision-making, particularly for fisheries, aquaculture, and the monitoring of environmentally sensitive events such as harmful algal blooms. However, accurately resolving coastal circulation and subsurface temperature variability remains challenging due to the influence of fine-scale atmospheric forcing, complex bathymetry, and sharp thermal gradients that are often inadequately represented in coarse-resolution models.

This study evaluates the skill of the SA-West forecasting system, which integrates regional ocean modelling with different atmospheric forcing products, and compares its performance against the baseline MERCATOR forecasts. The analysis focuses on key circulation features and subsurface temperature variability at selected coastal sites.

The aim is to assess whether the SA-West system provides improved predictive capability, particularly in nearshore regions, and to identify its strengths and limitations across different forecast lead times.

# 2. Methods and Materials

## Model description

All SOMISANA model developments utilize the Coastal and Regional Ocean COmmunity (CROCO) modelling system (https://www.croco-ocean.org/) that has been produced specifically to replicate regional, coastal and nearshore ocean processes. It utilizes terrain-following vertical coordinates (sigma-grid) and a free surface, which are crucial for accurately representing shallow coastal shelves and complex bathymetry. CROCO's core strength is its flexibility, including the ability to perform high-resolution simulations using two-way nesting (via AGRIF) and an optional non-hydrostatic solver for very fine-scale phenomena like strong vertical mixing or internal waves. Furthermore, it is designed for multi-model coupling with atmospheric, wave, sediment, and biogeochemistry models, providing the comprehensive physics required for holistic coastal ecosystem and operational forecasting studies. As a regional ocean modelling platform, CROCO models, both hindcast and forecast simulations, follow a downscaling approach, requiring the outputs from a global model to drive its lateral oceanic boundaries.

## Grid and bathymetry

The hindcast and forecast versions of the SOMISANA SA-West configuration V1.0 share the same grid. The model curvilinear grid has a variable horizontal grid resolution (1 – 2.5 km) which increases toward the coast (Figure 1). The model has 30 terrain-following (sigma) vertical levels that are built upon a model bathymetry which is interpolated from high resolution digital navigation charts provided by the South African National Hydrographic Office.

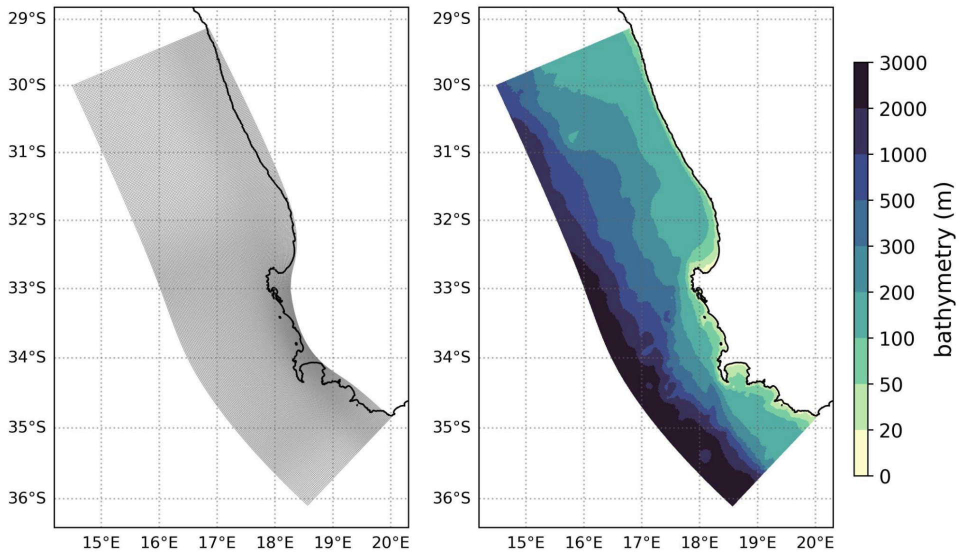


**Figure 1:** The SOMISANA SA-West V1.0 curvilinear grid configuration

The CROCO ocean model utilizes the Arakawa C-grid for its horizontal layout, a type of staggered grid that strategically separates where different physical variables are calculated. Instead of placing all variables at the center of a grid cell, CROCO places scalar quantities like temperature, salinity, and sea surface height at the cell's center, known as the rho-point. Conversely, the horizontal velocity components (u for east-west/zonal flow and v for north-south/meridional flow) are positioned at the respective faces of the cell, directly perpendicular to the flow they represent. This specific staggering is essential for the model's accuracy, as it allows for precise, second-order calculations of mass, momentum, and energy fluxes across the cell faces, ensuring key physical laws—like the conservation of energy and potential vorticity—are maintained throughout the simulation of regional and coastal ocean dynamics. This staggered grid means that three different sets of lat and lon coordinates are provided in the output files.

## Forcing datasets

Both the hindcast and forecast configurations require datasets from external data providers to force the model, as detailed in Table 1.

**Table 1**: Summary of datasets used to force the SOMISANA SA-West V1.0 hindcast and forecast simulations.

| | Hindcast | Forecast |
|---|---|---|
| Surface forcing | ERA5[1], WASA3[2] | GFS[3], SAWS[4] |
| Open boundary conditions | GLORYS12V1[5] | MERCATOR[6], HYCOM[7] |
| Tides | Not implemented | TPXO10[8] |

1. ~ 30 km resolution reanalysis product from ECMWF (Hersbach et al., 2020; Copernicus Climate Change Service, Climate Data Store, 2023;)
2. ~ 3 km resolution WRF Wind Atlas for South Africa dataset provided by the Climate Systems Analysis Group (CSAG) (Hahmann et al., 2021). The outputs of a 5 year simulation are not available via the SOMISANA catalogue, but can be made available upon request.
3. ~ 30 km resolution Global Forecast System from NOAA (https://www.nco.ncep.noaa.gov/pmb/products/gfs/)
4. 4.4 km resolution atmospheric forecast model run operationally by the South African Weather Service (SAWS)
5. ~9 km resolution global ocean reanalysis product produced by Mercator Ocean International (Lellouche et al., 2019; https://doi.org/10.48670/moi-00021)
6. ~9 km resolution global ocean analysis and forecast product produced by Mercator Ocean International (https://doi.org/10.48670/moi-00016)
7. ~9 km resolution global ocean analysis and forecast product produced by the HYCOM consortium (https://www.hycom.org/dataserver/gofs-3pt1/analysis)
8. ~4 km resolution global tidal constituents (https://www.tpxo.net/global/tpxo10-atlas)

Table 1 indicates that the forecast configuration employs two different open boundary forcing products as well as two different surface forcing products. This allows us to provide a four-member ensemble of forecasts: MERCATOR-GFS, MERCATOR-SAWS, HYCOM-GFS and the HYCOM-SAWS which serves two purposes - greater spread of potential forecast ocean states and increased redundancy in the operational system.

# Parameterisations

The hindcast and forecast versions of the SOMISANA SA-West configuration V1.0 share the same model parameterizations, as detailed in Table 2.

**Table 2**: Parameterisation details for the SOMISANA SA-West configuration V1.0 (capitalised keys in brackets refer to the CPP key names used in compiling the CROCO source code for this configuration)

| | |
|---|---|
| CROCO version | 1.3.1 |
| Horizontal grid | Curvilinear (128 x 446 grid points) |
| Vertical grid | Sigma (30 vertical levels) |
| Horizontal advection: momentum | 3rd-order upstream-biased (UV_HADV_UP3) |
| Horizontal advection: tracers | Rotated-Split 3rd-order upstream (TS_HADV_RSUP3) |
| Vertical advection: momentum | Spline vertical advection (UV_VADV_SPLINES) |
| Vertical advection: tracers | Spline vertical advection (TS_VADV_SPLINES) + adaptive semi-implicit vertical advection (VADV_ADAPT_IMP) |
| Horizontal mixing | Disabled - no explicit lateral viscosity or tracer diffusion |
| Vertical mixing | Generic Length Scale closure, k-epsilon (GLS_MIXING) |
| Barotropic open boundary conditions | Characteristic scheme for 2D momentum (OBC_M2CHARACT) |
| Baroclinic open boundary conditions | Orlanski radiation for 3D momentum (OBC_M3ORLANSKI) and tracers (OBC_TORLANSKI) |
| Air-sea fluxes | COARE 3.0 bulk formulae (BULK_FLUX) |
| Tides | Disabled in the hindcast, enabled in the forecast (TIDES) |
| Rivers | Disabled |

# Additional Datasets:

## DUACS

This dataset provides gridded daily and monthly mean global estimates of sea level based on satellite altimetry measurements. The sea level dataset is distributed by Copernicus Marine Environmental Monitoring System (CMEMS) and the details can be found following this link. The dataset includes estimates of sea level anomaly and absolute dynamic topography with a variety of other derived parameters (geostrophic currents, geostrophic current anomalies, eddy kinetic energy, etc.). We used the absolute dynamic topography for the comparison between our model and this dataset. The absolute dynamic topography is the sea surface height above the geoid computed as the sum of the sea level anomaly with the mean dynamic topography

### Underwater Temperature Measurements (UTR)

The Acoustic Tracking Array Platform (ATAP) is hosted by the South African Institute for Aquatic Biodiversity (SAIAB) and an official partner of the Canadian-based global Ocean Tracking Network (OTN). They monitor acoustically tagged marine animals using more than 285 moored acoustic receivers spanning ~2300 km of the southern African coastline, from St Helena Bay in the Western Cape, South Africa, to Santa Maria in southern Mozambique. More about ATAP can be found following this link. They provide hourly temperature measurements at their acoustic stations which are used opportunistically to determine the skill of our forecast system. In our skill analysis we used the UTR stations in Table 3, indicating their site, station name, location and depth deployed.

**Table 3:** Metadata of the Underwater Temperature Measurements (UTRs) used from the Acoustic Tracking Array Platform (ATAP).

| Site | Station | Latitude | Longitude | Bottom depth | Instrument Depth |
|---|---|---|---|---|---|
| St. Helena Bay | VKV001 | -32.7438 | 18.0358 | 12 | 10.5 |
| St. Helena Bay | VKV002 | -32.7653 | 18.08878 | 6 | 4.5 |
| St. Helena Bay | VKV003 | -32.7452 | 18.0779 | 12 | 10.5 |
| St. Helena Bay | VKV005 | -32.7398 | 18.11982 | 10 | 8.5 |
| St. Helena Bay | VKV006 | -32.7379 | 18.14772 | 5 | 3.5 |
| Cape Point | CP001 | -34.3355 | 18.48023 | 32 | 30.5 |
| Cape Point | CP002 | -34.3321 | 18.48548 | 50 | 48.5 |
| Cape Point | CP003 | -34.3289 | 18.49085 | 55 | 53.5 |
| Cape Hangklip | FB001 | -34.3555 | 18.8069 | 40 | 38.5 |
| Cape Hangklip | FB002 | -34.3567 | 18.8006 | 50 | 48.5 |
| Cape Hangklip | FB003 | -34.3578 | 18.79421 | 58 | 56.5 |

## Derived Parameters:

### Geostrophic Currents

The meridional and zonal geostrophic currents are defined as:

$$v_g = \frac{1}{f}\frac{\partial\eta}{\partial x} \text{ and}$$
$$u_g = -\frac{1}{f}\frac{\partial\eta}{\partial y},$$

where $v_g$ and $u_g$ are the respective meridional and zonal geostrophic components in $m\,s^{-1}$. $f$ denotes the Coriolis parameter and $\partial\eta$ denotes the meridional and zonal gradient of sea-surface height in the east-west ($\partial x$) and north-south ($\partial y$) directions. We chose to derive the geostrophic current from the DUACS dataset to ensure consistency when comparing our model forecasts.

## Statistical Parameters:

**Table 4:** Statistical Parameters used in the skill analysis. Description and formula used for each statistical parameter are provided.

| Parameter | Description | Formula |
|---|---|---|
| Bias | Computed by subtracting the observations ($X_{Obs}$) from the model ($X_{FC}$) which gives an indication that the model is either over/underestimating. | $bias = X_{FC} - X_{Obs}$ |
| Standard Deviation (SD) | The standard deviation (SD) is a measure of the amount of variation of the values of a variable about its average. A low/high SD indicates that the values tend to be close/spread to the average. The SD is calculated by computing the difference between each value ($x$) and the mean ($\bar{x}$), squaring those differences, averaging them by dividing by the total number of values ($N$), and then taking the square root of that result. | $SD = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}}$ |
| Correlation Coefficient ($r$) | Pearson's correlation coefficient ($r$) is used to quantify how closely the model predictions correspond to the measurements. Correlation ranges from -1 to | $r = \frac{\Sigma(X_{Obs} - \overline{X_{Obs}})(X_{FC} - \overline{X_{FC}})}{\sqrt{\Sigma(X_{Obs} - \overline{X_{Obs}})^2 \Sigma(X_{FC} - \overline{X_{FC}})^2}}$ |

| | | |
|---|---|---|
| | 1. Positive (Negative) values are indicative of a similar (opposite) relationship and zero is indicative of no relationship. The correlation coefficient is calculated by multiplying the deviations of the observations ($X_{Obs}$) and model forecasts ($X_{FC}$) values from their means (delineated by the overbar), summing these products, and dividing by the product of their SD. | |
| Centered Root Mean Square Error (CRMSE) | The Centered Root Mean Square Difference (RMSD) is used to measure the difference between the temperature measurements recorded by the UTRs ($X_{Obs}$) and the forecasted temperatures by the model ( $X_{FC}$) after their means have been removed (delineated by the overbar). By removing the means it ignores any seasonal or long term bias. N is the total number of forecasts. | $CRMSE = \sqrt{\frac{\Sigma\left[\left(X_{Obs}(t) - \overline{X_{Obs}}\right) - \left(X_{FC}(t) - \overline{X_{FC}}\right)\right]^2}{N}}$ |
| Skill Score (SS) | The SS is a measurement to quantify the model performance compared to the observations. We use a routine defined by Tonani et al. (2009) and Murphy et al. (1988).<br>The SS is the ratio between the CRMSE between the observations and the forecast model, called OF, and the CRMSE between the | $OF = CRMSE$<br><br>$OP = \sqrt{\frac{\Sigma\left[\left(X_{Obs}(t) - \overline{X_{Obs}}\right) - \left(X_{Obs}(t{=}d1) - \overline{X_{Obs}}\right)\right]^2}{N}}$<br><br>$SS = (1 - \frac{OF}{OP}) \times 100$ |

| | | |
|---|---|---|
| | observations and the persistence, called OP.<br>The difference between OF and OP is how the persistence is defined, which is $X_{Obs}(t = d1)$, where $d1$ is the observations at the forecast day 1.<br><br>We subtract the ratio between OF and OP from 1 and multiply by 100. The result is:<br>SS = 100% if OF is equal to zero - perfect forecast.<br>SS = 0 if OF is equal to OP - very poor forecast.<br>SS > 0 if the OF is less than OP which means that the forecast error is less than the persistence - gain of the forecast with respect to persistence. The contrary holds for SS<0. | |

# 3. Skill Analysis

The SA-West forecast system consists of four ensemble members based on the available datasets (Table 1). Each member includes one ocean dataset (lateral boundary conditions) and one atmospheric dataset (surface forcing).

The naming convention for the ensemble members reflects the forcing products used. For example, the SA-West MERCATOR-GFS forecast model refers to the South African West Coast forecast model that uses MERCATOR at its lateral open boundaries and GFS for surface forcing.

The GFS-forced members provide a 5-day forecast, whereas the SAWS-forced members provide a 3-day forecast. All ensemble members derive their tidal forcing from TPXO10.

A skill analysis of the MERCATOR-GFS and MERCATOR-SAWS forecasts is provided at the time of writing this.

## 3.1. Hydrodynamical Flow Field

The SA-West forecast runs (MERCATOR-GFS and MERCATOR-SAWS) are compared with DUACS satellite altimetry to assess how well the forecast products capture large-scale hydrodynamic features along the west coast.
Figure 2 presents instantaneous maps of sea-surface height (red contours) and surface-derived geostrophic currents (black arrows), comparing satellite altimetry (left column) with the MERCATOR-GFS (centre column) and MERCATOR-SAWS (right column) forced runs.

The Benguela Jet and the recirculation within St Helena Bay are key dynamical features that must be accurately represented in the SA-West forecast models due to their environmental significance in the region. The Benguela Jet is characterised by an equatorward flow inshore of the 500 m isobath (thin black contour), between Cape Point in the south and Cape Columbine in the north (Figure 2a).

In the forecast simulations, the Benguela Jet and the recirculation within St Helena Bay are well captured up to day 3 of the forecast runs (centre and right columns, Figure 2). By day 5 of the MERCATOR-GFS forecast, the Benguela Jet intensifies, restricting the recirculation within St Helena Bay to the north.

This analysis indicates that the large-scale flow is well represented in the forecast models, especially with the shorter forecasts. It is suggested that the accuracy starts to diminish for the longer forecast periods.

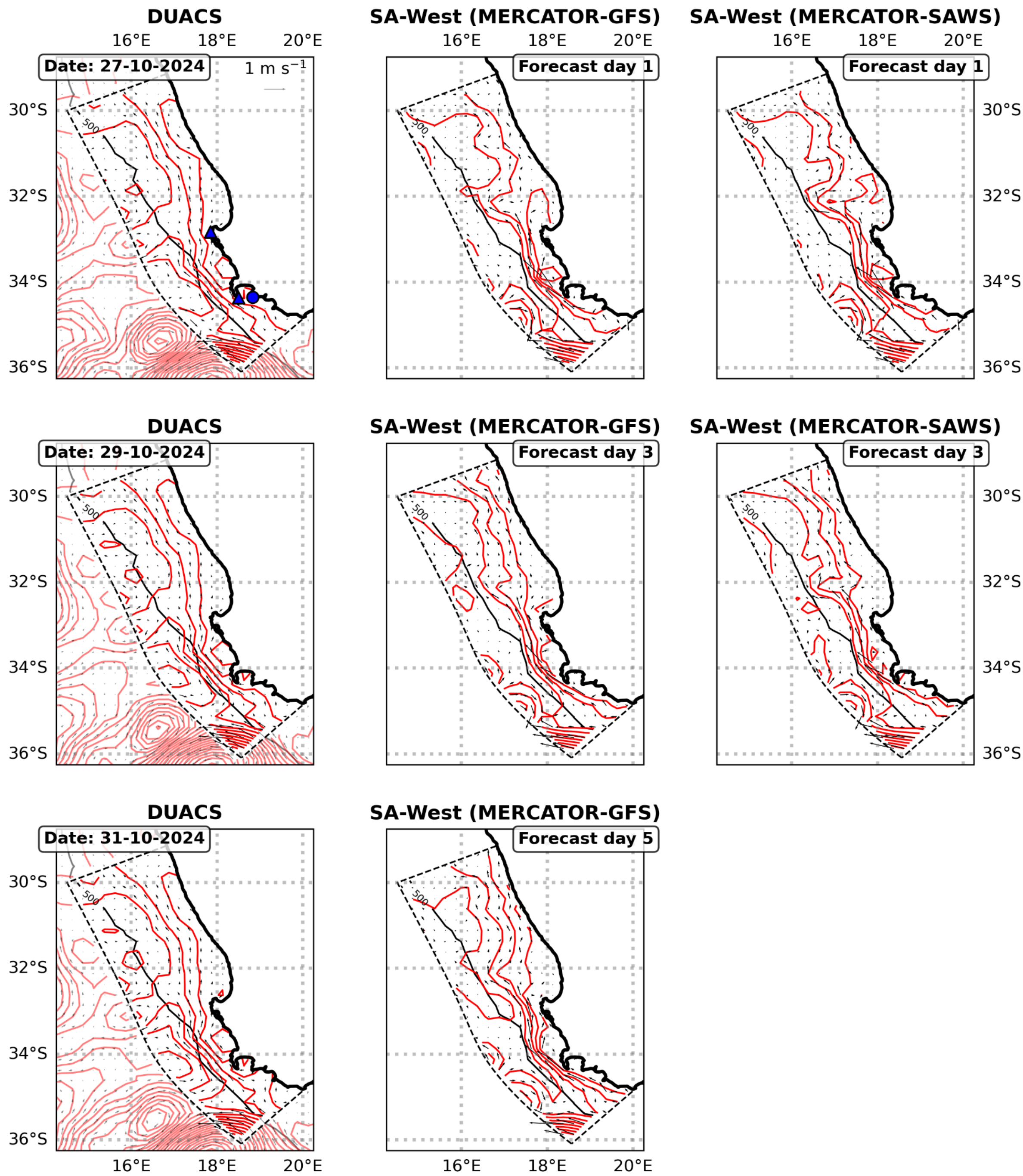


**Figure 2:** Snapshots of sea-surface height (red contours) and surface derived geostrophic currents (black arrows) comparing DUACS satellite altimetry (left column) to the SA-West Forecast models. The center and right columns are the MERCATOR-GFS and MERCATOR-SAWS forecast runs, respectively. The forecast runs were initialised on 16 August 2024. In the top left figure, triangles show the position of Cape Columbine in the north and Cape Point in the south and the circle the position of the Cape Hangklip UTR.

## 3.2. Subsurface Temperatures

Due to the close proximity of the UTRs, red dots in Figure 3, we assume similar dynamics at each site and averaged the UTRs within each cluster, namely: St. Helena Bay, Cape Point and Cape Hanglip (red boxes in Figure 3). We compare the clustered temperature time-series to both the SA-West and MERCATOR forecasts time-series. The nearest grid point in the forecast models was used to compare to the spatial mean position of the clustered time-series positions.

All temperature time-series were averaged daily. In the time-series analysis we compare the skill of SA-West and MERCATOR forecast models ability to predict subsurface temperatures using UTR measurements as the reference. The aim is to show that the SA-West provides an improved prediction of the subsurface temperatures compared to the MERCATOR forecasts.

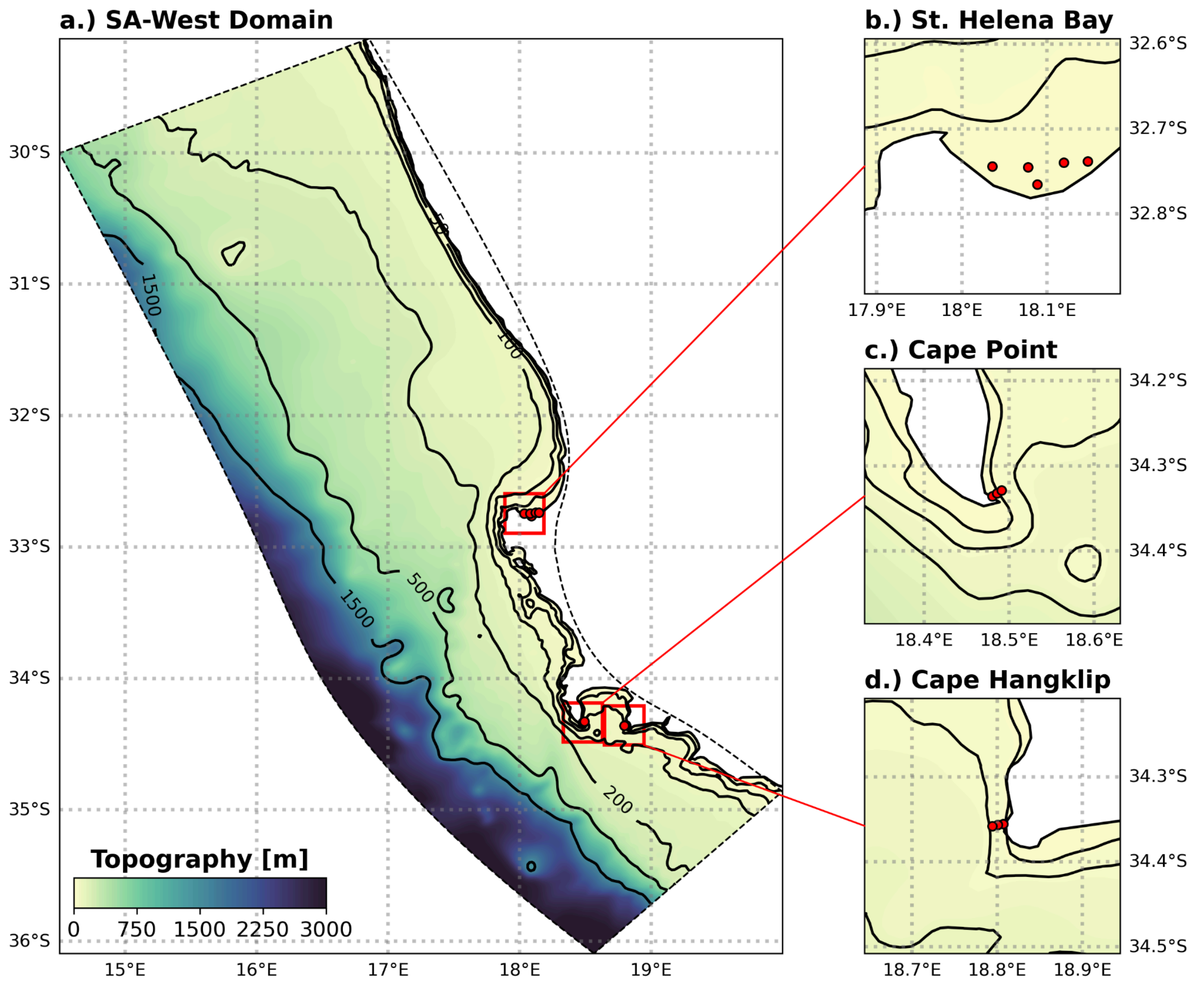


**Figure 3:** (a.) Topographic map of the SA-West forecast model domain with black contours representing specific depths. Depths are defined at 50 m, 100 m, 200 m, 500 m and 1500 m. Underwater Temperature Recorders (UTR) are indicated with red dots. UTRs are clustered together in St. Helena Bay (b.), Cape Point and (c.) Cape Hangklip in red boxes which are zoomed in on the adjacent figures.

The subsurface temperatures at each of the clusters (St. Helena Bay, Cape Point and Cape Hangklip) are analysed in detail in the following subsections.

### 3.2.1. St. Helena Bay

The mean UTR temperatures at St. Helena Bay (Figures 4 and 5), Cape Point (Figures 5 and 6), and Cape Hangklip (Figures 6 and 7) are compared with temperatures extracted from the SA-West (top) and MERCATOR (bottom) forecasts. The UTRs are shown as a black time series, whereas the forecasts are shown in different colours, with each colour representing a different forecast day. The bias between the model forecast days and the UTRs is displayed in the bottom left of each time series.

The Taylor diagram adjacent to each time series summarises the standard deviation (SD), correlation coefficient (r), and centered root mean square difference (RMSD) between the forecast models and the UTRs. The points closest to the reference point (solid black line) are the most accurate. The distance from the origin (0,0) represents the standard deviation (variability) of the model. The angle from the horizontal x-axis represents the correlation coefficient between the model and the observations. The distance between a model point and the reference point represents the RMSD.

The MERCATOR and MERCATOR-GFS forecasts do not provide much statistical confidence at the St. Helena Bay station due to low correlations ($r < 0.1$) and high RMSE ($RMSE > 0.8$) values (Figures 4). The correlations ($r > 0.48$) and RMSE ($RMSE < 0.8$) are better represented in the MERCATOR-SAWS run (Figures 5). It is noteworthy that both SA-West forecast models produce a lower temperature bias and follow the temperature pattern more closely. This is reflected in the elevated correlations in the MERCATOR–SAWS run, which highlights the importance of the finer-scale wind patterns provided by SAWS and resolved in the SA-West model.

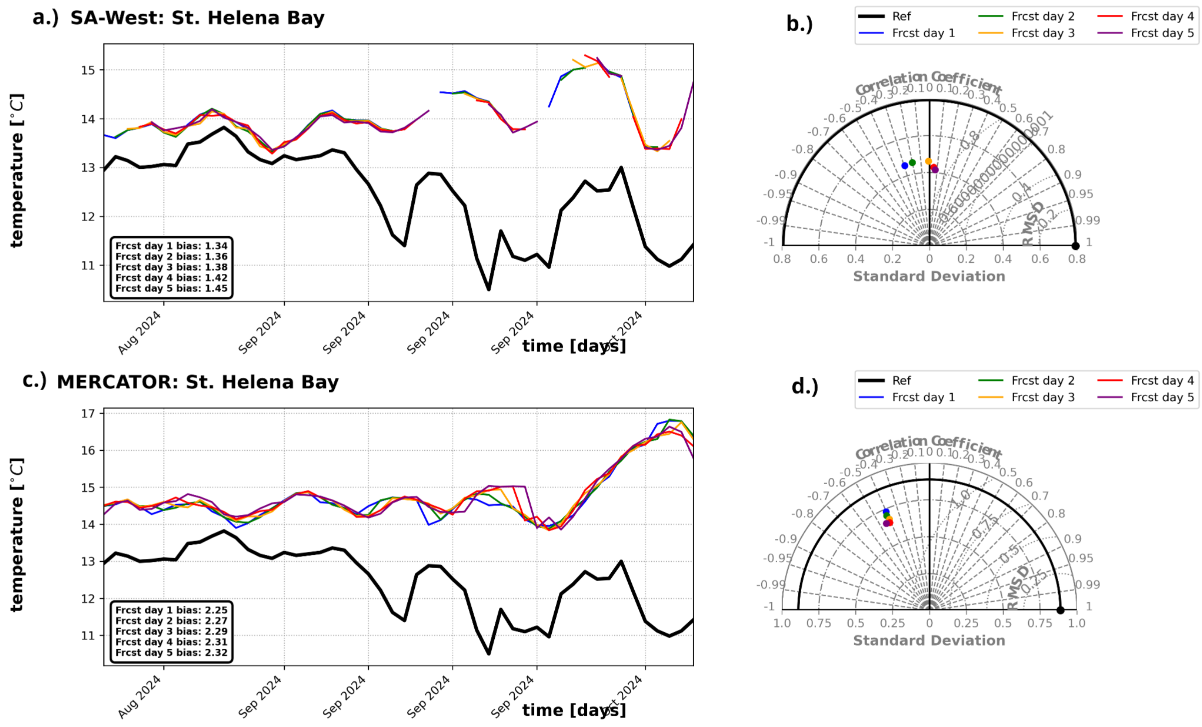


**Figure 4:** Time-series of the (a.) SA-West (MERCATOR-GFS) and (c.) MERCATOR modelled temperatures, in colour, compared to the UTR measurements, black, of the Cape Hangklip cluster. The different colours represent the forecast days: blue = forecast day 1, green = forecast day 2, yellow = forecast day 3, red = forecast day 4, purple = forecast day 5. Average bias between the model and the UTR measurements are indicated on the bottom left of the time-series. The standard deviation, correlation coefficient and the centered Root Mean Square Difference (RMSD) are represented on the Taylor diagram adjacent to the time-series.

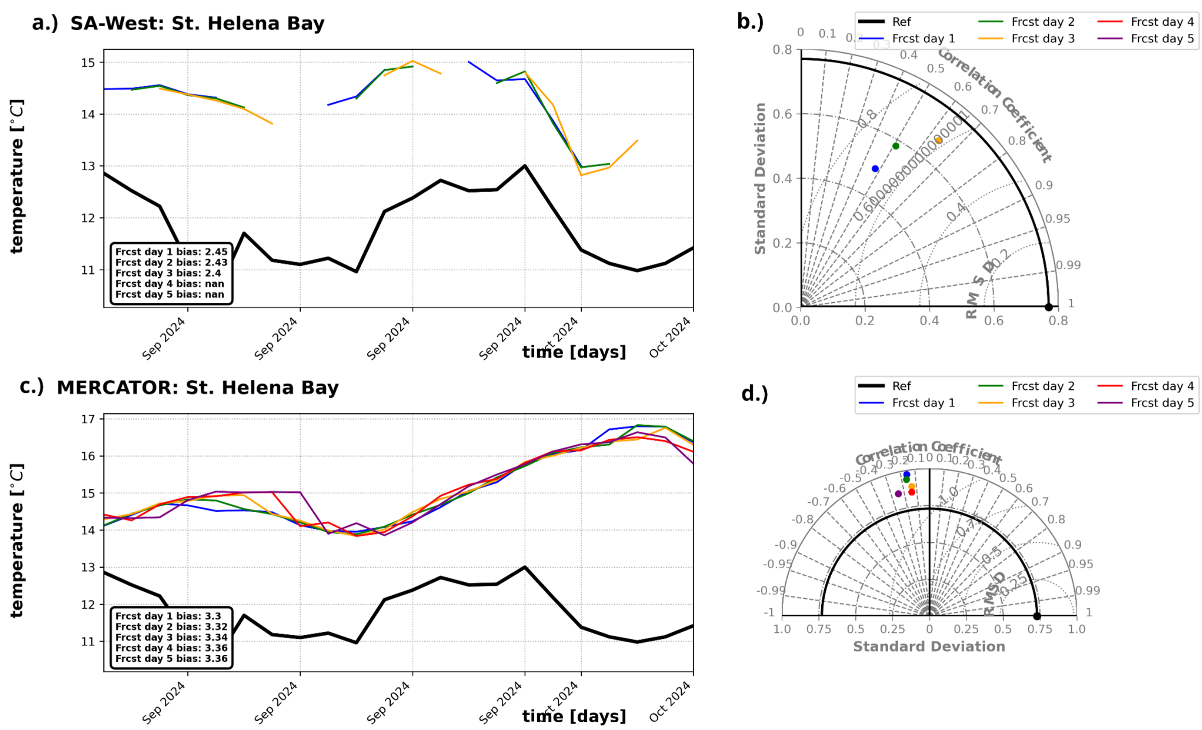

**Figure 5:** Time-series of the (a.) SA-West (MERCATOR-SAWS) and (c.) MERCATOR modelled temperatures, in colour, compared to the UTR measurements, black, of the St. Helena Bay cluster. The different colours represent the forecast days: blue = forecast day 1, green = forecast day 2, yellow = forecast day 3, red = forecast day 4, purple = forecast day 5. Average bias between the model and the UTR measurements are indicated on the bottom left of the time-series. The standard deviation, correlation coefficient and the centered Root Mean Square Difference (RMSD) are represented on the Taylor diagram adjacent to the time-series.

### 3.2.1. Cape Point

At Cape Point, both SA-West and MERCATOR accurately follow the subsurface temperature measurements, confirmed by the low bias (-0.8 < bias < 0.2), low RMSD values (RMSD < 1.1) and high correlations (r > 0.4, Figure 6). It is noteworthy to point out that MERCATOR does tend to retain its accuracy throughout the 5 day forecast, whereas SA-West degrades after 3 days - indicative of the clustering of the daily points in the Taylor diagram (Figure 6). However, the first 3 days showed better performance by SA-West.

Even though the bias (bias < -0.25) and correlations (r < 0.22) are slightly poorer for the MERCATOR-SAWS run in Figure 7, it is promising to note a similar degree of variability in the SD (points close to the black reference line). This is largely underestimated by MERCATOR which again highlights the finer scale dynamics captured in the MARCATOR-SAWS forecasts.

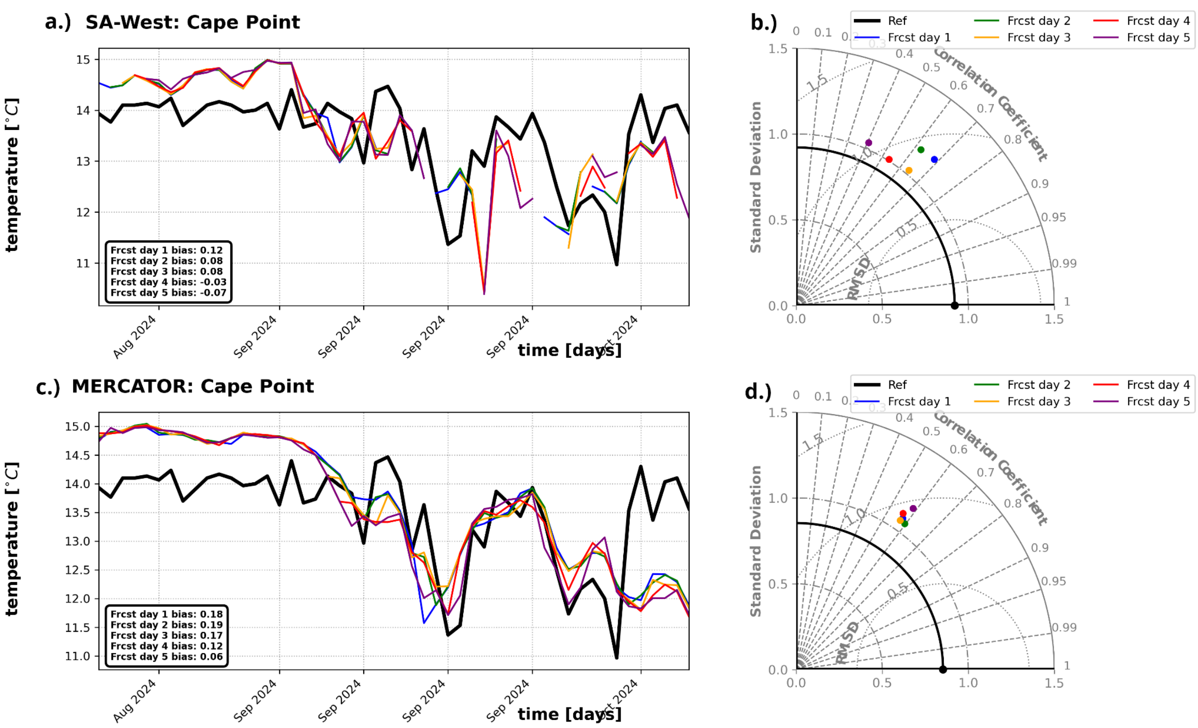


**Figure 6:** Time-series of the (a.) SA-West (MERCATOR-GFS) and (c.) MERCATOR modelled temperatures, in colour, compared to the UTR measurements, black, of the Cape Hangklip cluster. The different colours represent the forecast days: blue = forecast day 1, green = forecast day 2, yellow = forecast day 3, red = forecast day 4, purple = forecast day 5. Average bias between the model and the UTR measurements are indicated on the bottom left of the time-series. The standard

deviation, correlation coefficient and the centered Root Mean Square Difference (RMSD) are represented on the Taylor diagram adjacent to the time-series.

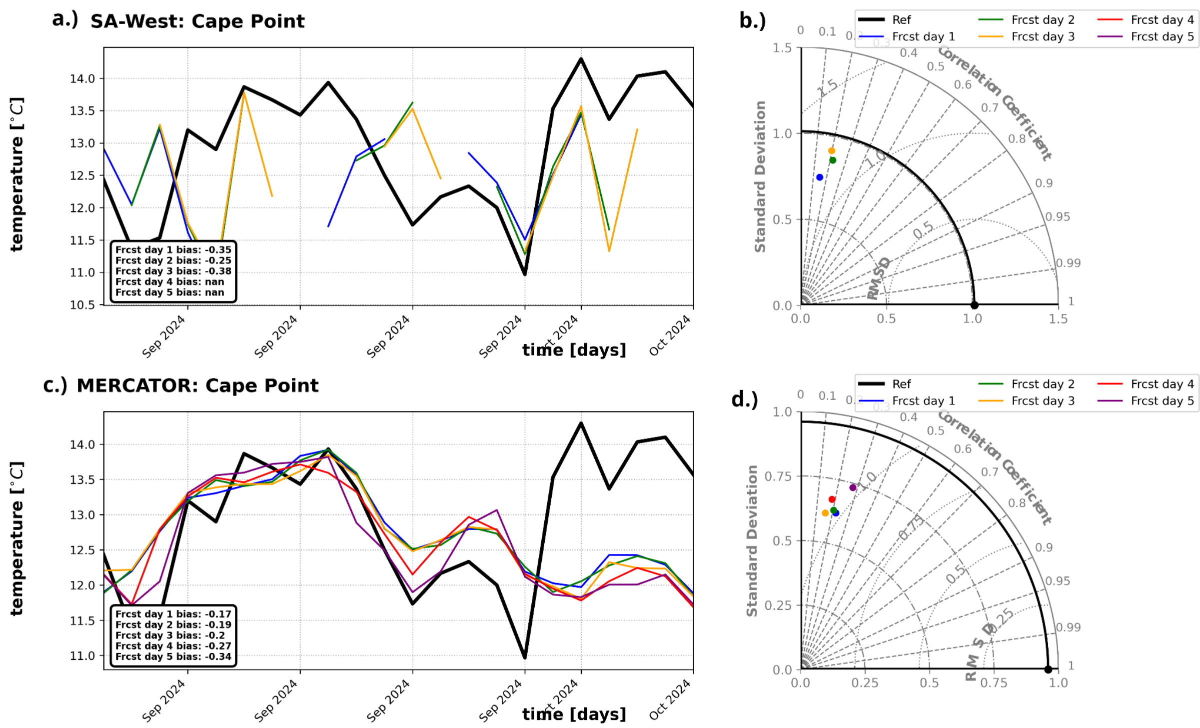


**Figure 7:** Time-series of the (a.) SA-West (MERCATOR-SAWS) and (c.) MERCATOR modelled temperatures, in colour, compared to the UTR measurements, black, of the Cape Point cluster. The different colours represent the forecast days: blue = forecast day 1, green = forecast day 2, yellow = forecast day 3, red = forecast day 4, purple = forecast day 5. Average bias between the model and the UTR measurements are indicated on the bottom left of the time-series. The standard deviation, correlation coefficient and the centered Root Mean Square Difference (RMSD) are represented on the Taylor diagram adjacent to the time-series.

### 3.2.3. Cape Hangklip

The MERCATOR-GFS run performs particularly well at Cape Hangklip compared to MERCATOR (Figures 8). This is confirmed by the low bias (bias < 0.59), low RMSD values (RMSD < 0.77), and high correlations (r > 0.79). In addition, the MERCATOR-GFS time series closely follows that of the UTR, and the points on the Taylor plot are grouped together and lie close to the reference line, suggesting that the forecast can retain its accuracy over the entire 5-day and depicts a similar degree of variability. MERCATOR, on the other hand, does not resolve the temperature in this cluster well, resulting in high bias (bias > 1.69), high RMSD (RMSD > 1.25) values, and low correlations (r = < 0.15).

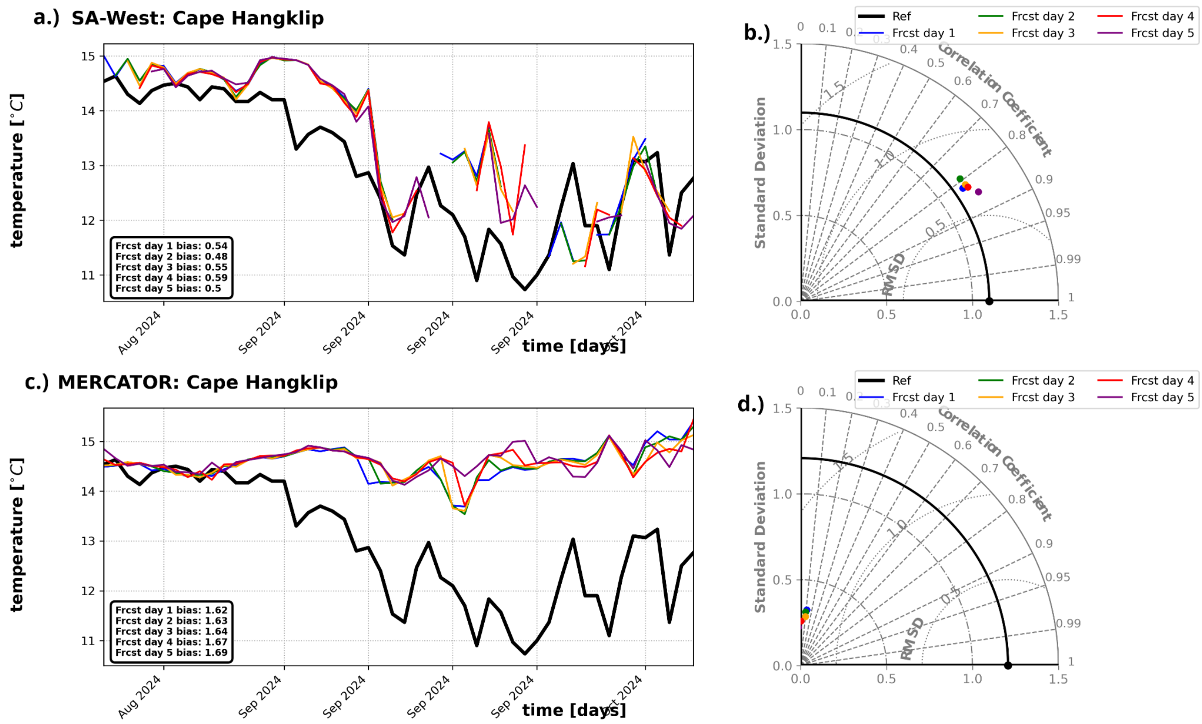


**Figure 8:** Time-series of the (a.) SA-West (MERCATOR-GFS) and (c.) MERCATOR modelled temperatures, in colour, compared to the UTR measurements, black, of the Cape Hangklip cluster. The different colours represent the forecast days: blue = forecast day 1, green = forecast day 2, yellow = forecast day 3, red = forecast day 4, purple = forecast day 5. Average bias between the model and the UTR measurements are indicated on the bottom left of the time-series. The standard deviation, correlation coefficient and the centered Root Mean Square Difference (RMSD) are represented on the Taylor diagram adjacent to the time-series.

MERCATOR-SAWS does capture the subsurface temperature accurately (-0.04 < bias < 0.16) and its variability - close proximity to the reference line in the Taylor diagram. The correlations are initially high, but do drop significantly over the 3 day forecast which suggest that the forecast struggles to retain its accuracy.

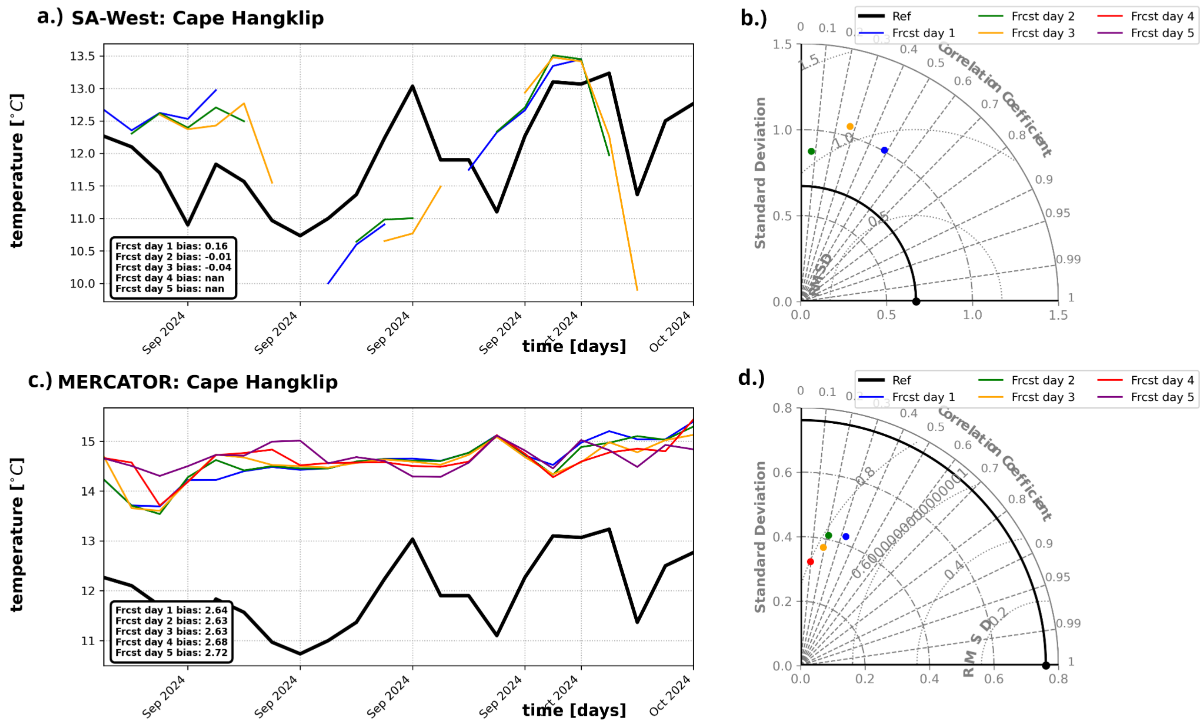


**Figure 9:** Time-series of the (a.) SA-West (MERCATOR-SAWS) and (c.) MERCATOR modelled temperatures, in colour, compared to the UTR measurements, black, of the Cape Hangklip cluster. The different colours represent the forecast days: blue = forecast day 1, green = forecast day 2, yellow = forecast day 3, red = forecast day 4, purple = forecast day 5. Average bias between the model and the UTR measurements are indicated on the bottom left of the time-series. The standard deviation, correlation coefficient and the centered Root Mean Square Difference (RMSD) are represented on the Taylor diagram adjacent to the time-series.

## 3.3. Skill Score

The skill of the SA-West forecast models (blue) are compared to the MERCATOR forecasts (red) to determine if the SA-West models outperforms the MERCATOR forecasts (Figure 10). It should be noted that Tonani et al. (2009) computed their skill score from the reanalysis product of their forecast models (GLORYS12, distributed by CMEMS) which yielded higher skill scores.

In our case, we computed the skill score from the UTR measurements, which is very strict on the models and resulted in very low skill scores (SS < 0). Technically this is indicative of poor forecasts, however, in this comparison we aim to show that we can outperform the MERCATOR forecasts.

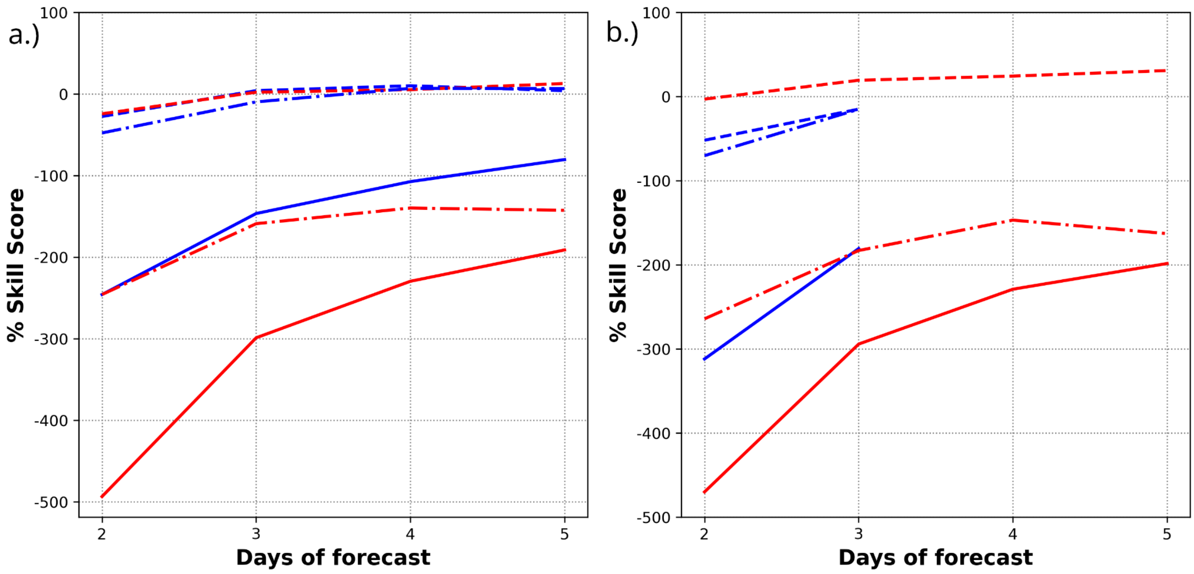


**Figure 10:** Skill Scores of the SA-West (blue) and MERCATOR (red) forecasts. MERCATOR-GFS and MERCATOR-SAWS forecasts are shown on the left and right figures, respectively. The clusters are represented with the various linestyles, where: solid = St. Helena Bay, dash = Cape Point, and dashdot = Cape Hangklip.

The MERCATOR–GFS forecast performs better at the St. Helena Bay (solid) and Cape Hangklip (dash-dot) clusters (blue > red in Figure 10a), while showing similar performance to MERCATOR at Cape Point (dashed). Although MERCATOR-GFS performs better between days 3 and 4 at Cape Point, MERCATOR slightly outperforms the SA-West model by day 5.

Similarly, the MERCATOR-SAWS forecast performs better at St. Helena Bay and Cape Hangklip, but at Cape Point, MERCATOR outperforms MERCATOR–SAWS (Figure 10b).

# 4. Summary and Concluding Remarks

The SA-West forecast system comprises four ensemble members that combine different ocean (MERCATOR) and atmospheric (GFS or SAWS) forcing products, with forecast horizons of 5 days (GFS) and 3 days (SAWS). A skill analysis was conducted for the MERCATOR-GFS and MERCATOR-SAWS configurations.

Comparison with satellite altimetry shows that both SA-West configurations realistically capture key large-scale hydrodynamic features, including the Benguela Jet and St. Helena Bay recirculation, particularly within the first 3 days of the forecast. Forecast skill degrades at longer lead times, with increased intensity of the Benguela Jet and reduced circulation evident by day 5.

Subsurface temperature validation against UTR observations indicates spatial variability in model performance:

- **St. Helena Bay:** Both SA-West and MERCATOR show low skill (low correlations, high RMSD), although SA-West exhibits reduced bias and better representation of temporal variability, especially in the SAWS-forced run.
- **Cape Point:** Both models perform well, with low bias and high correlations. SA-West shows improved performance in the first 3 days, while MERCATOR maintains skill over the full 5-day period.
- **Cape Hangklip:** SA-West outperforms MERCATOR, with consistently lower errors, higher correlations, and better representation of variability.

Skill score analysis (based on UTR observations) yields generally low absolute scores for all models but demonstrates relative improvements. SA-West configurations outperform MERCATOR at St. Helena Bay and Cape Hangklip, while MERCATOR performs similarly or better at Cape Point, particularly at longer lead times.

The SA-West forecast system demonstrates clear added value over the baseline MERCATOR forecast system, particularly in regions influenced by finer-scale coastal dynamics, such as Cape Hangklip. The inclusion of higher-resolution atmospheric forcing (SAWS) improves the representation of variability and reduces bias, highlighting the importance of local wind forcing.

Forecast skill is highest within the first 3 days, after which performance declines, especially for SA-West configurations. While MERCATOR retains more consistent skill at longer lead times, it generally underperforms in resolving coastal variability.

Overall, the SA-West system provides improved short-term predictive capability for subsurface temperatures and coastal circulation along the west coast. Future improvements should focus on including longer forecasts periods to ensure robust statistics.

# 5. Reference List

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly weather review*, *116*(12), 2417-2424.

Tonani, M., Pinardi, N., Fratianni, C., Pistoia, J., Dobricic, S., Pensieri, S., De Alfonso, M. & Nittis, K. (2009). Mediterranean Forecasting System: forecast and analysis assessment through skill scores. *Ocean Science*, *5*(4), 649-660.