# Appendix

The details of the comparison between different model configurations and AQUA_MODIS data have been presented in this section. The comparison with reanalysis data is also included.

## 1. AQUA_MODIS

All models' configuration sea surface temperature (SST) have been compared with AQUA_MODIS data (Figure 1). The data is monthly data for September 2020. After interpolating the temperature of the ROMS to the AQUA_MODIS for each configuration, the error of the model has been calculated as SST (AQUA_MODIS) - SST (ROMS). For all configurations, the model underestimates temperature in most of the Aegean Sea and Adriatic Sea but overestimates in other parts of the eastern Mediterranean Sea.

For i1 (Figure 2a) configuration, the model error is 1 to 1.5°C in the eastern part of the Ionian Sea, but it is –1°C for most parts of the Ionian Sea. In the south of the Ionian Sea, the error is +1°C. In the Adriatic Sea, in the eastern part, the model underestimates by roughly 0.5–1°C, but in the western part of the southern basin, the model overestimates by 0.5°C. In the Aegean Sea, in the north, the model underestimates by roughly –1°C, but in the eastern part, it overestimates by 0.5–1°C.

In i2 (Figure 2b) configuration, in the Adriatic, the model underestimates slightly by 0.5–1°C in most parts of the Adriatic. In the Aegean Sea, the model underestimates SST by 0.5–1°C in most parts, which is a better result compared to the i1 configuration. In the Ionian and Levantine Seas, the model overestimates by 0.5–1°C, which is also a better result compared to i1. In the i3 (Figure 2c) configuration, the model presents slightly better results compared to i1 and i2. However, in the Levantine Sea, near 28°E and 35°N, the error is at its maximum. In the southern basin of the Ionian Sea, i3 presents slightly better results. In the Aegean Sea, i3 also shows some improvement.

For the h1 (Figure 2d), the model underestimates temperature by –1°C in the eastern part of the southern Adriatic and the eastern part of the northern Ionian Sea. In the Aegean Sea, the model underestimates by –0.5 to –1°C. In other parts of the eastern Mediterranean Sea, the model overestimates SST by 0.5–1°C. In h2 (Figure 2e), in the Adriatic Sea, the model underestimates SST by 0.5–1.5°C. In the Aegean Sea, the model underestimates SST by 0.5–1°C in the east but overestimates it in the west. Therefore, compared to h1, there is no remarkable improvement in h2.

For v1 (Figure 2f), the model underestimates SST by 0.5–1°C in the Adriatic Sea, with the maximum error along the east coast. In the Aegean Sea, the model underestimates SST by 0.5–1°C, while in other parts, it overestimates SST by roughly 0.5–1°C. However, in one area of the Levantine Sea, there is an error of 2°C. In v2 (Figure 2g), compared to v1, the model's underestimation error increases by 0.5°C in the Adriatic Sea. In the Ionian Sea, the model's error is 0.1–0.2°C higher compared to v1. Therefore, it seems that the model presents better results in the v1 configuration.

The performance of the seven configurations (i1, i2, i3, h1, h2, v1, and v2) was statistically evaluated in comparison with AQUA_MODIS data (Table 1). The sensitivity of the model under different setups and configurations was not remarkable. Among the configurations, h1 and i2 exhibit the best overall performance, with h1 achieving the highest correlation coefficient ($r$ = 0.88) and the lowest RMSE (0.85°C), indicating strong agreement with observed data and minimal error. Similarly, i2 records the lowest Mean Absolute Error (MAE = 0.63°C) and Mean Absolute Bias (MAB = 0.63°C), making it highly accurate in predicting temperature values. Both configurations have minimal biases, further supporting their reliability. On the other hand, i3 stands out for its nearly unbiased predictions (Mean Bias = –0.01°C) but slightly lags in correlation ($r$ = 0.88).

In contrast, v2 delivers the performance with the highest MAE (0.72°C), RMSE (1.01°C), and the lowest correlation ($r$ = 0.84), suggesting greater differences between model predictions and observations. Similarly, v1 and h2 exhibit higher errors and biases compared to the top-performing configurations. These results suggest that, while all configurations maintain reasonably strong correlations with AQUA_MODIS data, the sensitivity of the model to different setups and configurations is not remarkable.

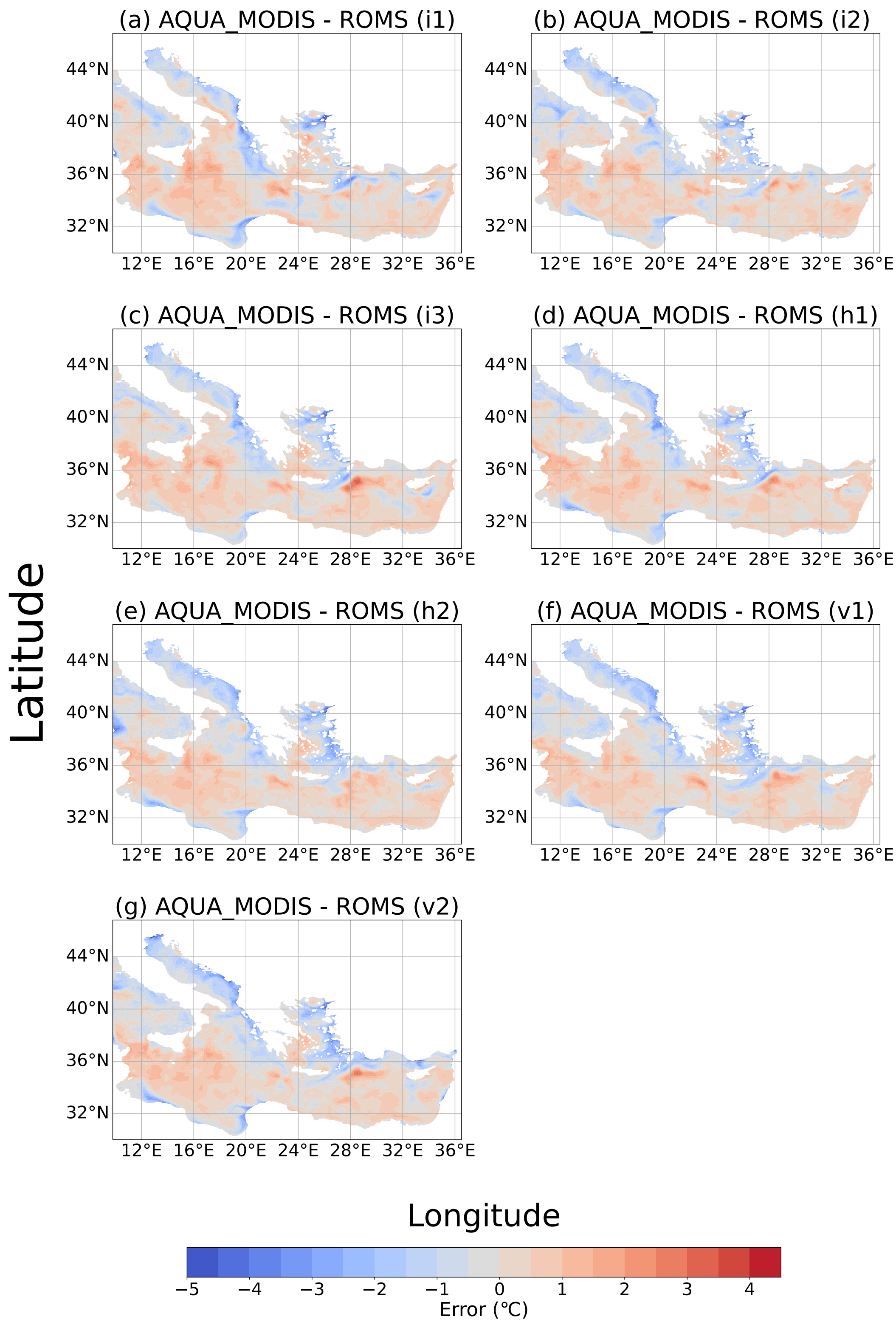

(a) AQUA_MODIS - ROMS (i1)
(b) AQUA_MODIS - ROMS (i2)
(c) AQUA_MODIS - ROMS (i3)
(d) AQUA_MODIS - ROMS (h1)
(e) AQUA_MODIS - ROMS (h2)
(f) AQUA_MODIS - ROMS (v1)
(g) AQUA_MODIS - ROMS (v2)
44°N
40°N
36°N
32°N
12°E 16°E 20°E 24°E 28°E 32°E 36°E
Latitude
Longitude
−5 −4 −3 −2 −1 0 1 2 3 4
Error (°C)

**Figure 1:** Model temperature errors for different configurations compared to the AQUA_MODIS dataset. Errors were calculated as Error = AQUA_MODIS - ROMS.

**Table 1:** Performance of seven configurations (i1, i2, i3, h1, h2, v1, v2) evaluated against AQUA_MODIS. Metrics: MAE (Mean Absolute Error) measures average error magnitude, RMSE (Root Mean Square Error) quantifies error dispersion, MB (Mean Bias) indicates an average signed error, and MAB (Mean Absolute Bias) reflects unsigned average error magnitude. SST (AQUA_MODIS) - SST(ROMS) is used for calculations.

| Configuration | MAE (ºC) | RMSE(ºC) | MB(ºC) | MAB(ºC) | Correlation |
|---|---|---|---|---|---|
| i1 | 0.66 | 0.89 | 0.03 | 0.66 | 0.85 |
| i2 | 0.63 | 0.84 | 0.08 | 0.63 | 0.89 |
| i3 | 0.64 | 0.85 | -0.01 | 0.64 | 0.88 |
| h1 | 0.63 | 0.85 | 0.02 | 0.63 | 0.88 |
| h2 | 0.66 | 0.89 | 0.13 | 0.66 | 0.88 |
| v1 | 0.67 | 0.89 | 0.17 | 0.67 | 0.88 |
| v2 | 0.72 | 1.01 | 0.17 | 0.72 | 0.84 |

## 2. Reanalysis Data

For better assessment and verification of the model, we compared it with reanalysis data for both monthly averages and daily values. As obtaining satellite data daily for the entire domain is challenging, the daily comparison is conducted only using reanalysis data. In the monthly average analysis, we focus on September (4.2.1), when Ianos occurred. For the daily analysis (4.2.2), we specifically examine 17 September, the duration of Ianos. In the following sections, we discuss the rationale behind selecting this particular day in more detail.

## 3. Monthly

For i1(Figure 3a), SST results from ROMS show an overestimate of 0.5 to 1°C in most study areas, except for the south and northeast of the Ionian Sea and the north of the Levantine Sea, where the error reaches up to 1.5°C. i2 (Figure 3b) presents better results, with noticeable improvements in the Adriatic Sea and the Aegean Sea. In the Ionian Sea, the error is reduced by 0.1 to 0.2°C compared to i1. In i3 (Figure 3c), better results are seen in the Adriatic Sea compared to i1 and i2, but in the north of the Levantine Sea, the error increases, reaching up to 2.5°C. In the Ionian Sea, the i3 results are better overall.

For h1 (Figure 3d), the Adriatic Sea model typically underestimates temperatures by 0.5 to 1°C in the eastern part but overestimates by 0.2°C in the western part. In the north and east of the Aegean Sea, there is an overestimation of

0.5 to 1°C, but in the west, there is an underestimation of 1 to 1.5°C. In the Ionian Sea, the model overestimates by 0.5 to 1°C. In h2 (Figure 3e), improvements are seen in the Ionian Sea, but errors increase in the Adriatic and Aegean Seas. For v1 (Figure 3f), the model shows an underestimation of up to 1°C in the eastern Adriatic but an overestimation in the western Adriatic. In the Aegean Sea, the model underestimates by 0.5 to 1°C, while in the Levantine Sea, there is an overestimation of 0.5 to 1°C. In v2 (Figure 3g), errors increase in the northern and eastern Adriatic, especially near coastal waters, but are less remarkable in other regions.

The performance of the seven configurations (i1, i2, i3, h1, h2, v1, and v2) was statistically evaluated against reanalysis data (Table 2). Among these, simulations v1 and i2 emerged as the most reliable, with v1 achieving the lowest Mean Absolute Error (MAE) and Mean Absolute Bias (MAB) at 0.52°C and a high correlation coefficient ($r$=0.92). Similarly, i2 demonstrated the lowest Root Mean Squared Error (RMSE) at 0.69°C, with a strong correlation ($r$=0.92). Both configurations showed minimal biases, indicating their close alignment with the reanalysis data. In contrast, v2 displayed the weakest performance, with the highest RMSE (0.85°C) and the lowest correlation (r=0.88), suggesting it deviates more from the observed trends.

Other simulations, including i3, h1, and h2, performed comparably well with almost similar metrics across MAE, RMSE, and correlation coefficients ($r$=0.91 or 0.92). These models exhibited small biases, with h2 standing out for having the smallest mean bias (MB = –0.02°C). However, the slightly higher MAE and RMSE metrics for these configurations placed them just behind the top-performing v1 and i2. Overall, v1 and i2 demonstrated the best balance of low errors and strong correlation, making them the most accurate simulations in this analysis.

Regarding salinity, in i1 (Figure 4a), the model underestimates salinity by 0.5–0.8 PSU in most areas but overestimates it in the northern Aegean, Adriatic, and western Ionian Sea. In i2 (Figure 4b), results are improved, particularly in the Ionian Sea. In i3 (Figure 4c), the model shows better results in the middle and southern Adriatic and parts of the Ionian and Levantine Seas. For h1 (Figure 4d), errors are negligible in most of the Adriatic and the center and south of the Aegean Sea, but in the Levantine Sea, the model underestimates salinity by 0.6 to 0.7 PSU. In most of the Ionian Sea, errors are negligible. In h2 (Figure 4e), errors increase in most areas. For v1 (Figure 4f), the model overestimates salinity by 0.5 to 0.6 PSU in the Adriatic, with errors increasing to 1.5 PSU in the northern Aegean Sea. In the Levantine Sea and northern Ionian Sea, the model underestimates salinity by 0.5 to 0.7 PSU. In v2 (Figure 4g), there are no remarkable differences compared to v1.

The performance of seven different ROM simulations (i1, i2, i3, h1, h2, v1, and v2) was evaluated based on key error metrics in comparison to reanalysis salinity data (Table 3). i3 stands out as the best performer overall, with the lowest Mean Absolute Error (MAE) of 0.39 PSU and Root Mean Squared Error (RMSE) of 0.48 PSU, as well as a strong correlation coefficient of 0.77. This indicates that i3 provides the most accurate and reliable salinity predictions with minimal bias. On the other hand, i1 is the weakest model, with a high MAE of 0.59 PSU, RMSE of 0.67 PSU, and a very low correlation of just 0.20, suggesting that it cannot calculate the salinity variations effectively.

The remaining simulations (i2, h1, h2, v1, and v2) show relatively strong performances, with i2 achieving the highest correlation (0.79) but having a larger bias compared to other models. h1, h2, v1, and v2 perform similarly in terms of error metrics, with MAE and RMSE values ranging from 0.40 to 0.44 PSU and correlation values between 0.73 and 0.76. While these models perform well in capturing salinity, they show slightly higher errors compared to i3. Overall, the analysis reveals that i3 and h1 are the most accurate, while i1 needs improvements to match the performance of the other simulations.

In circulation comparisons, the worst results are in i1 (Figure 5a). However, i2 (Figure 5b) shows improvements, particularly in the Adriatic gyres, Aegean Sea, and eastern Levantine Sea. In i2, the model underestimates current speed by 0.1 m/s in the Levantine Sea and the center of the Ionian Sea. In i3 (Figure 5c), better compatibility of circulation is observed. In h1 (Figure 5d), there are improvements in the structure of eddies in the Ionian Sea, which are more notable in h2 (Figure 5e). This indicates that decreasing grid spacing leads to better representation of eddy and gyre structures. For increased vertical levels, v1 (Figure 5f) and v2 (Figure 5g), there are no remarkable differences, though stronger currents are seen in some regions, such as the Levantine Sea.

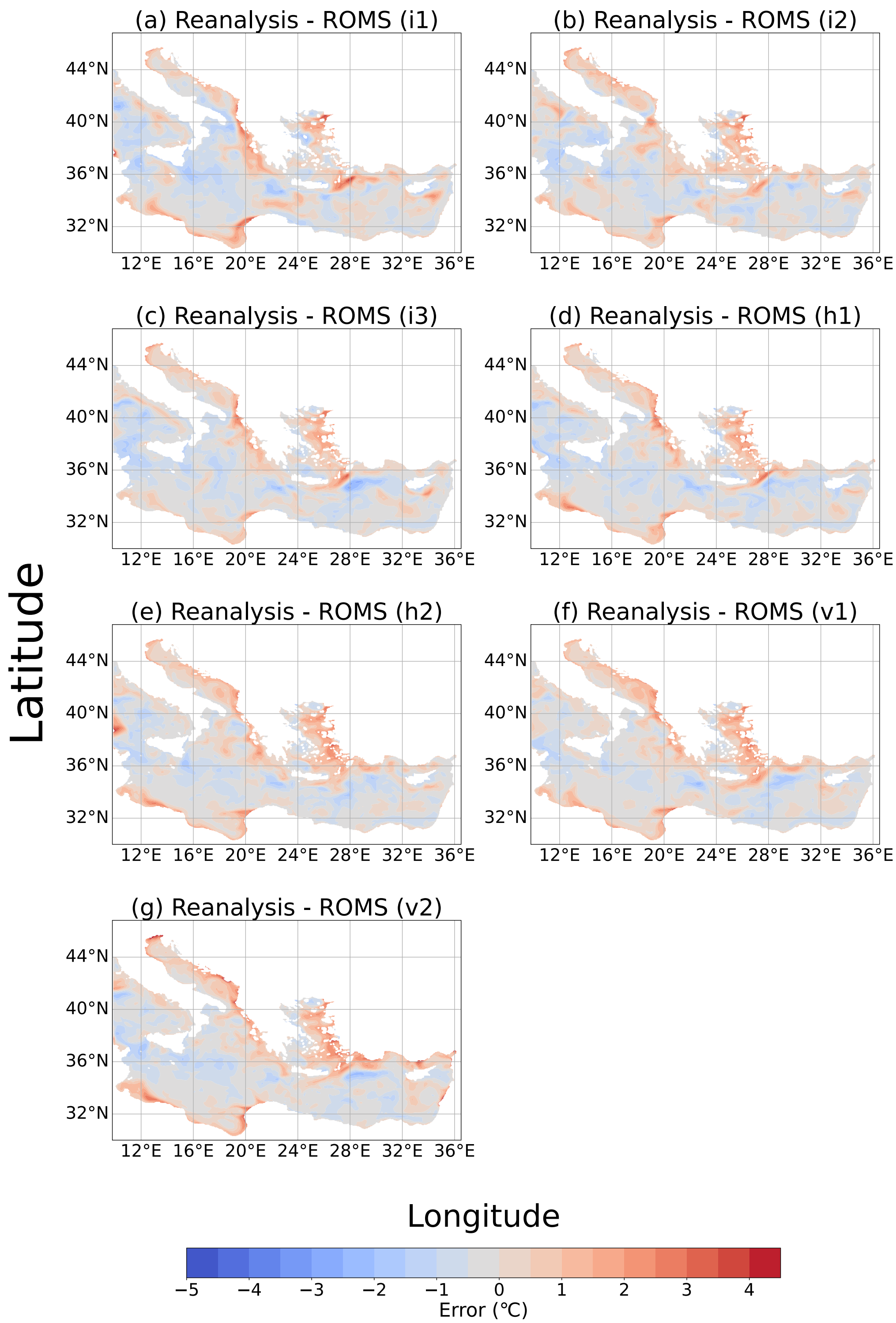


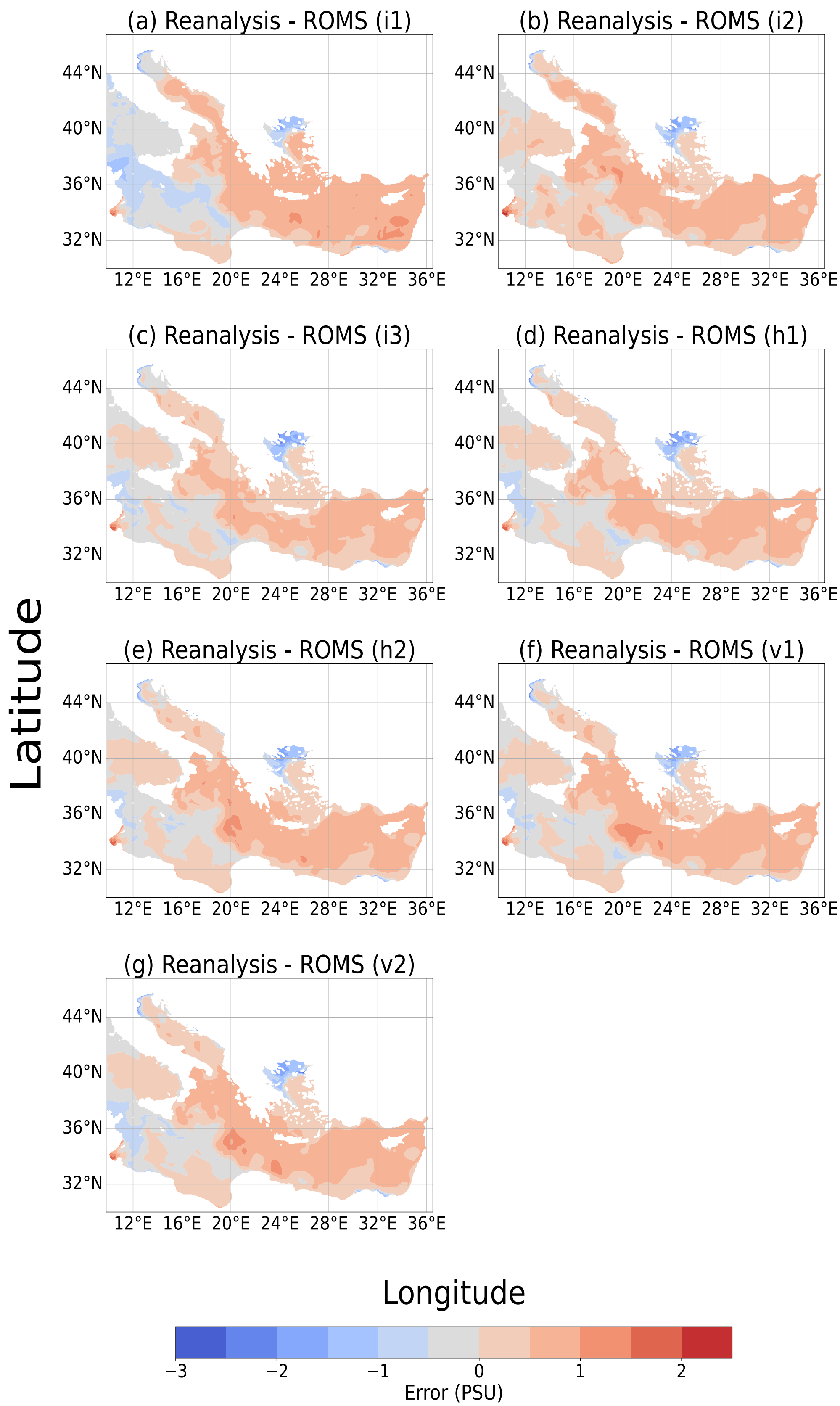

**Figure 2:** Model temperature errors for different configurations compared to the reanalysis dataset in September, based on monthly average data. Errors were calculated as Error = ROMS - reanalysis.

**Table 2:** Performance metrics for seven configurations compared to reanalysis for temperature. Key metrics used for assessment are defined as follows: MAE: Mean Absolute Error, representing the average magnitude of errors. RMSE: Root Mean Square Error, the square root of the average squared errors. MB: Mean Bias, indicating the average error magnitude, including its sign. MAB: Mean Absolute Bias, reflecting the average magnitude of errors without considering direction.

| Configuration | MAE (ºC) | RMSE (ºC) | MB (ºC) | MAB (ºC) | Correlation |
|---|---|---|---|---|---|
| i1 | 0.59 | 0.78 | -0.11 | 0.59 | 0.88 |
| i2 | 0.53 | 0.68 | -0.06 | 0.53 | 0.92 |
| i3 | 0.54 | -0.70 | -0.15 | 0.54 | 0.92 |
| h1 | 0.55 | 0.72 | -0.12 | 0.55 | 0.91 |
| h2 | 0.54 | 0.72 | -0.02 | 0.54 | 0.92 |
| v1 | 0.52 | 0.70 | 0.02 | 0.52 | 0.92 |
| v2 | 0.60 | 0.85 | 0.03 | 0.60 | 0.88 |

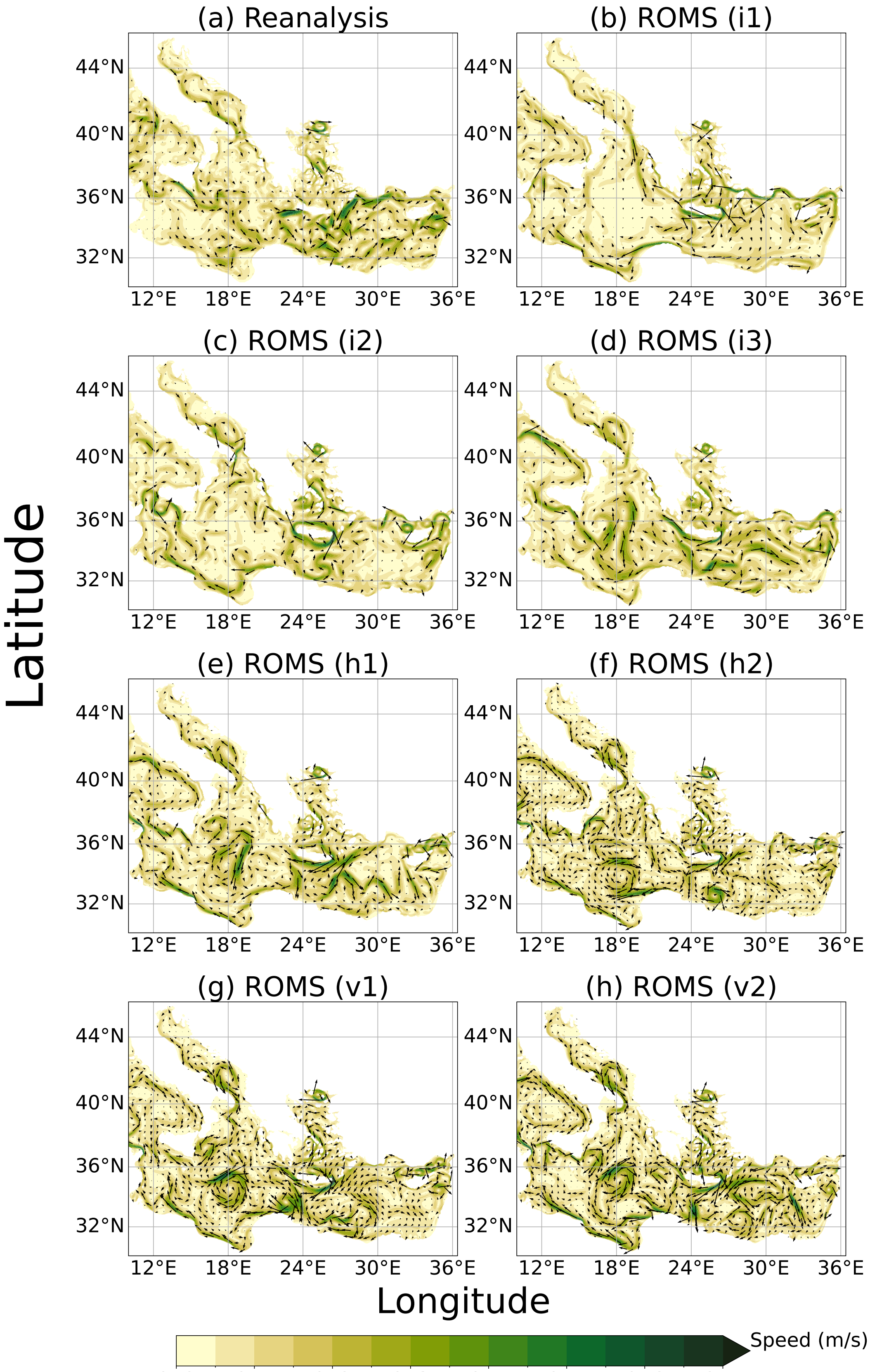
**Table 3:** Performance Metrics for seven configurations compared to reanalysis. Key metrics used for assessment are defined as follows: MAE: Mean Absolute Error, representing the average magnitude of errors. RMSE: Root Mean Square Error, the square root of the average squared errors. MB: Mean Bias, indicating the average error magnitude, including its sign. MAB: Mean Absolute Bias, reflecting the average magnitude of errors without considering direction.

| Configuration | MAE (PSU) | RMSE (PSU) | MB (PSU) | MAB (PSU) | Correlation |
|---|---|---|---|---|---|
| i1 | 0.59 | 0.67 | 0.24 | 0.59 | 0.20 |
| i2 | 0.45 | 0.53 | 0.35 | 0.45 | 0.79 |
| i3 | 0.39 | 0.48 | 0.20 | 0.39 | 0.77 |
| h1 | 0.40 | 0.49 | 0.20 | 0.40 | 0.75 |
| h2 | 0.43 | 0.52 | 0.27 | 0.43 | 0.76 |
| v1 | 0.43 | 0.53 | 0.26 | 0.43 | 0.74 |
| v2 | 0.44 | 0.53 | 0.25 | 0.44 | 0.73 |



(a) Reanalysis - ROMS (i1)
(b) Reanalysis - ROMS (i2)
(c) Reanalysis - ROMS (i3)
(d) Reanalysis - ROMS (h1)
(e) Reanalysis - ROMS (h2)
(f) Reanalysis - ROMS (v1)
(g) Reanalysis - ROMS (v2)
44°N
40°N
36°N
32°N
12°E 16°E 20°E 24°E 28°E 32°E 36°E
Latitude
Longitude
−3 −2 −1 0 1 2
Error (PSU)

**Figure 3:** Model salinity errors for different configurations compared to the reanalysis dataset in September, using monthly average data. Errors were computed as Error = ROMS - reanalysis.



(a) Reanalysis
(b) ROMS (i1)
(c) ROMS (i2)
(d) ROMS (i3)
(e) ROMS (h1)
(f) ROMS (h2)
(g) ROMS (v1)
(h) ROMS (v2)
44°N
40°N
36°N
32°N
12°E
18°E
24°E
30°E
36°E
Latitude
Longitude
Speed (m/s)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

**Figure 4:** Surface circulation of models with different configurations compared to the reanalysis dataset in September, based on monthly average data.

## 4. Daily

A comparison of seven ROMS model configurations (i1, i2, i3, h1, h2, v1, and v2) highlights a common issue: all models tend to overestimate temperatures near the coast, with warm biases reaching up to +3°C (Figure 5). However, the extent and distribution of these errors vary. The i-series (i1, i2, i3) show similar error patterns, with strong coastal overestimations and scattered offshore underestimations. The h-series (h1, h2) demonstrate some improvements by reducing offshore errors, though coastal biases remain. The v-series (v1, v2) perform best, particularly v2, particularly below the Ianos, which minimizes coastal warm biases while maintaining a balanced distribution of errors.

In terms of salinity, i3 emerges as the most accurate configuration, with the lowest Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and bias, along with a strong correlation (0.77) (Figure 6). Conversely, i1 has the highest errors and lowest correlation (0.15), making it the least reliable. i2 achieves the highest correlation (0.79) but still exhibits a higher bias than i3. The v2 and h1 configurations show moderate performance. A comparison of daily (September 17, 2020) and monthly (September 2020) models indicates consistent performance across timescales, with i3 maintaining the best accuracy. However, i1 shows a drop in correlation on the daily scale, indicating reduced reliability for short-term predictions.

For temperature, h1 is the top-performing configuration, with the lowest MAE (0.60°C), RMSE (0.82°C), and a strong correlation (0.87). Other configurations (i1, i2, i3, h2, v1, v2) have similar MAE values (0.66–0.67°C) but slightly lower accuracy than h1. Negative biases are most pronounced in i2 (–0.10°C) and v2 (–0.13°C), while h1 slightly overestimates (bias: 0.08°C). The i3 configuration maintains a near-zero bias (–0.02°C), balancing over- and underestimations. The monthly model (September 2020) consistently outperforms the daily model (17 September 2020), showing lower errors and higher correlation. The best monthly configuration is v1, while h1 performs best in the daily model. Overall, v2 is the best model for temperature predictions, while i3 is the most accurate for salinity. Monthly models offer more stable and reliable forecasts compared to daily simulations.

For circulation on 17 September, (Figure 7) the reanalysis dataset shows strong currents, with speeds up to 0.7 m/s and well-developed eddies. Among the ROMS simulations, h1, v1, and v2 best match the reanalysis, accurately capturing key circulation features and current strengths. h2 and v2 show moderate agreement but slightly underestimate speeds, while i1 and i3 deviate the most, with weaker currents and less-defined eddies. Overall, h1, v1, and v2 provide the most realistic Mediterranean circulation compared to the reanalysis.

Consequently, we return to our first scientific question: how do factors such as horizontal grid spacing vertical levels, and initial conditions affect model accuracy? In general, for temperature and salinity at the surface (except for i1), the sensitivity of the model to horizontal and vertical levels is not remarkable. However, in circulation, lower grid spacing leads to better results. For the rest of this paper, we focus on the v2 configuration, as it incorporates the best initial conditions, the lowest grid spacing, and the highest vertical level. Since Ianos has a remarkable impact on surface dynamics, understanding eddy and gyre structures benefits from using a 3 km grid spacing, which better captures these features. Furthermore, as we try to study the effects of Ianos at levels near the surface, 50 vertical levels work better to capture its details.

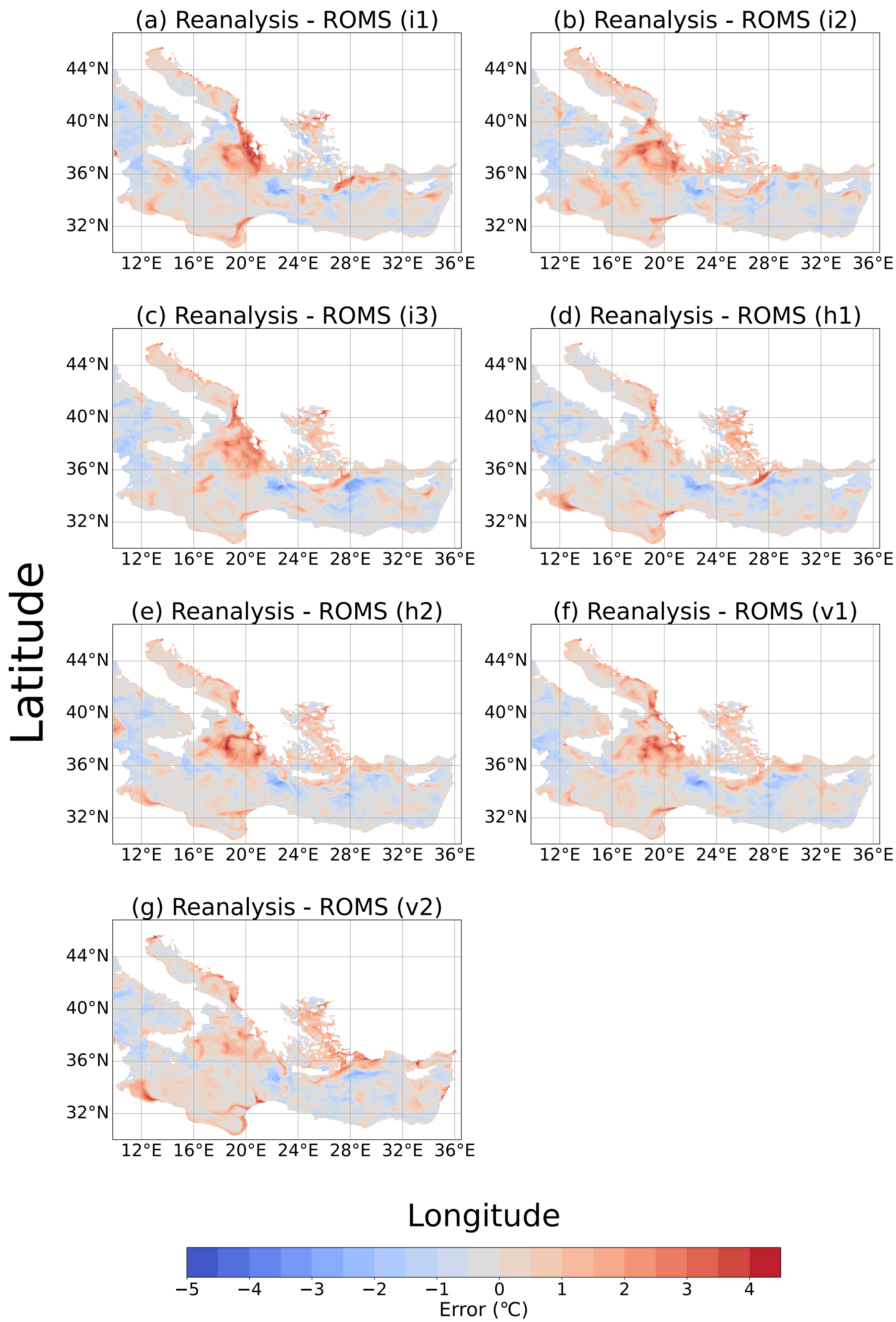



**Figure 5:** Model temperature errors for different configurations compared to the reanalysis dataset on 17 September, based on daily average data. Errors were calculated as Error = ROMS - reanalysis.

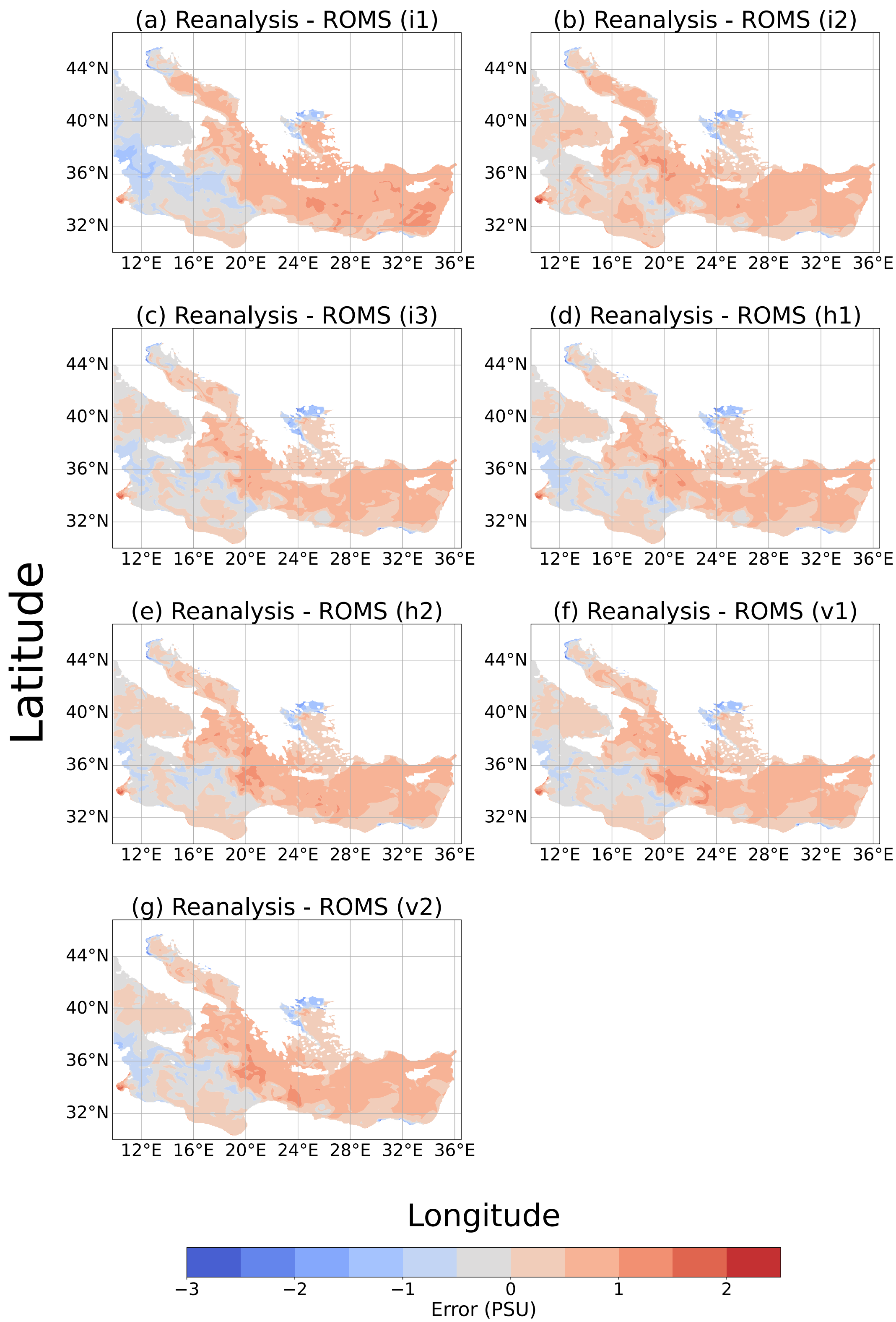



**Figure 6:** Model salinity errors for different configurations compared to the reanalysis dataset on 17 September, using daily average data. Errors were computed as Error = ROMS - reanalysis.

**Table 4:** Performance Metrics for seven configurations compared to reanalysis for 17 September. Key metrics used for assessment are defined as follows: MAE: Mean Absolute Error, representing the average magnitude of errors. RMSE: Root Mean Square Error, the square root of the average squared errors. MB: Mean Bias, indicating the average error magnitude, including its sign. MAB: Mean Absolute Bias, reflecting the average magnitude of errors without considering direction.

| Configuration | MAE (PSU) | RMSE (PSU) | MB (PSU) | MAB (PSU) | Correlation |
|---|---|---|---|---|---|
| i1 | 0.60 | 0.68 | 0.23 | 0.60 | 0.15 |
| i2 | 0.45 | 0.54 | 0.36 | 0.45 | 0.79 |
| i3 | 0.40 | 0.49 | 0.20 | 0.40 | 0.77 |
| h1 | 0.41 | 0.50 | 0.20 | 0.41 | 0.75 |
| h2 | 0.44 | 0.53 | 0.27 | 0.44 | 0.77 |
| v1 | 0.44 | 0.54 | 0.26 | 0.44 | 0.75 |
| v2 | 0.45 | 0.54 | 0.25 | 0.45 | 0.73 |

**Table 5:** Performance metrics for seven configurations compared to reanalysis for temperature for 17 September. Key metrics used for assessment are defined as follows: MAE: Mean Absolute Error, representing the average magnitude of errors. RMSE: Root Mean Square Error, the square root of the average squared errors. MB: Mean Bias, indicating the average error magnitude, including its sign. MAB: Mean Absolute Bias, reflecting the average magnitude of errors without considering direction.

| Configuration | MAE (ºC) | RMSE (ºC) | MB (ºC) | MAB (ºC) | Correlation |
|---|---|---|---|---|---|
| i1 | 0.67 | 0.96 | -0.04 | 0.67 | 0.83 |
| i2 | 0.67 | 0.91 | -0.10 | 0.67 | 0.86 |
| i3 | 0.66 | 0.91 | -0.02 | 0.66 | 0.85 |
| h1 | 0.60 | 0.82 | 0.08 | 0.60 | 0.87 |
| h2 | 0.66 | 0.90 | -0.04 | 0.66 | 0.86 |
| v1 | 0.64 | 0.90 | -0.09 | 0.64 | 0.86 |
| v2 | 0.66 | 0.95 | -0.13 | 0.66 | 0.84 |

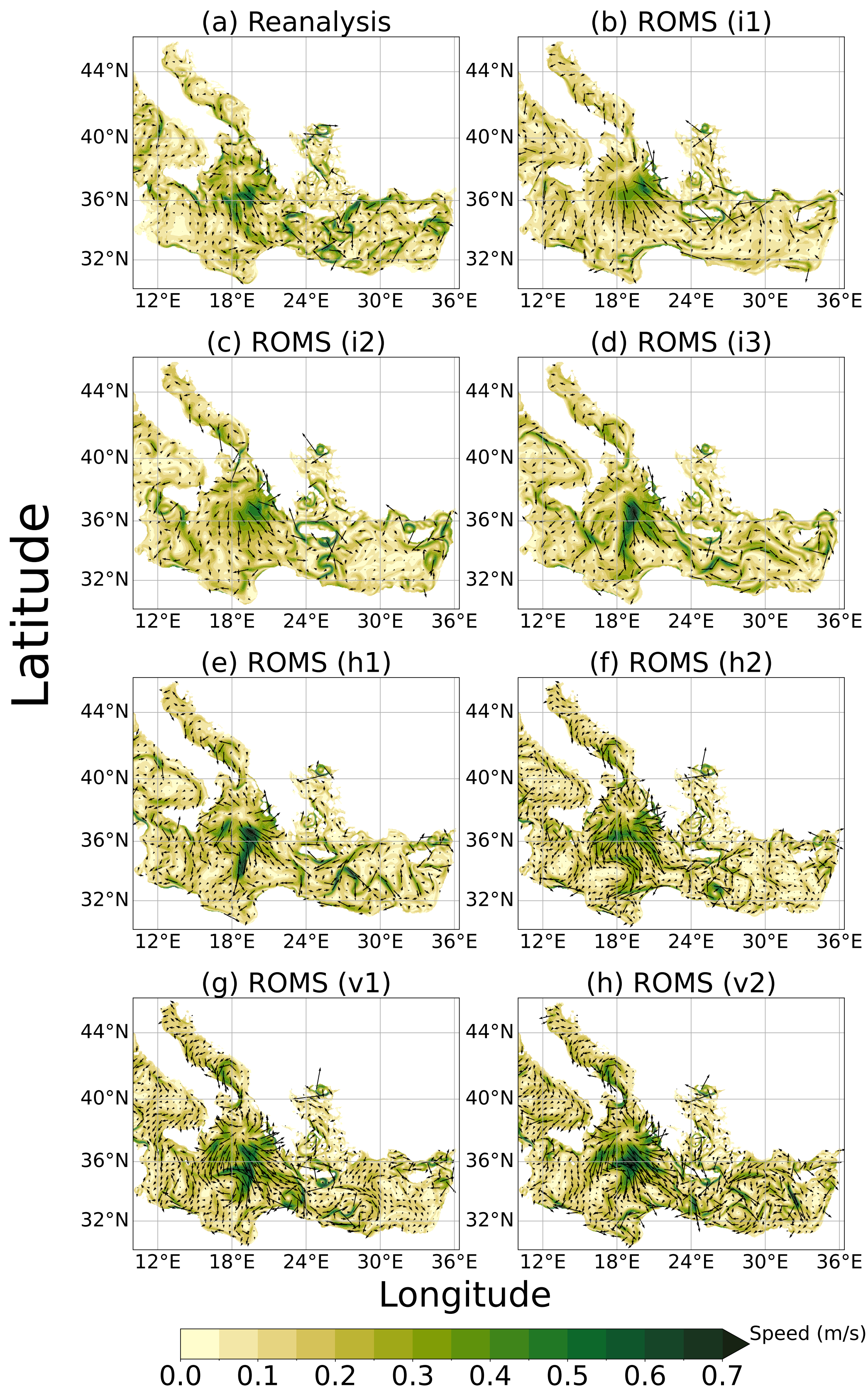

(a) Reanalysis
(b) ROMS (i1)
(c) ROMS (i2)
(d) ROMS (i3)
(e) ROMS (h1)
(f) ROMS (h2)
(g) ROMS (v1)
(h) ROMS (v2)
44°N
40°N
36°N
32°N
12°E
18°E
24°E
30°E
36°E
Latitude
Longitude
Speed (m/s)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

**Figure 7:** Surface circulation of models with different configurations compared to the reanalysis dataset in 17 September, based on daily average data.