James Lewis

UNIVERSITY OF EXETER

Assessment

Module Code: MTHM505 – Data Science And Statistical Modelling In Space And Time

---

Declaration of AI Assistance

I have used OpenAI's ChatGPT tool in creating this report.
AI-supported/AI-integrated use is permitted in this assessment. I acknowledge the following uses of GenAI tools in this assessment:

- Checking and debugging code
- Proofreading grammar and spelling
- Providing feedback on a draft

I declare that I have referenced use of GenAI outputs within my assessment in line with the University referencing guidelines.

Table of contents

1 Sea Surface Temperature Modelling

1.1 Part A: Cleaning and Spatial Overview

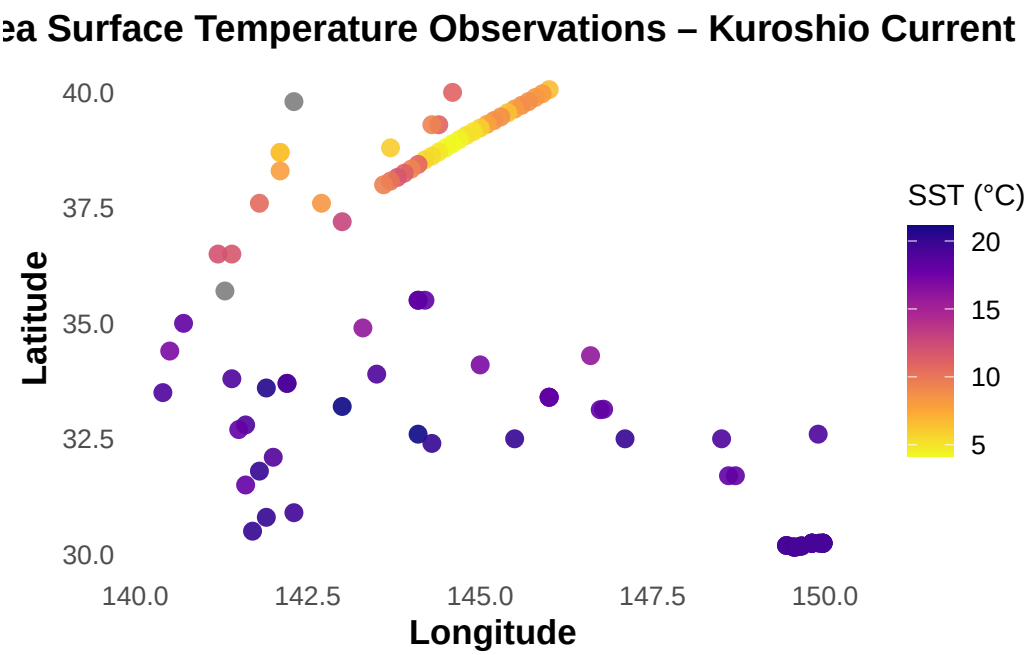


Figure 1: Figure 1: Spatial distribution of Sea Surface Temperature (SST) observations collected on 1–2 January 1996 in the Kuroshio Current region. Each point represents an individual measurement; colour denotes temperature, with warmer SSTs concentrated in the north-east band.

The dataset `kuroshio100.csv` contains 100 sea surface temperature (SST) observations from January 1996, recorded along the Kuroshio current system. Initial data inspection revealed three rows with missing spatial coordinates (`lon` or `lat`), which were removed to ensure compatibility with spatial modelling functions such as `as.geodata()`.

This resulted in **97 complete observations**, covering a broad range of longitudes and latitudes in the western Pacific Ocean. These values were retained for further exploratory and model-based analysis.

Figure 1 confirms the dataset captures a wide latitudinal spread and a broad SST range. Warmer values were observed to the south and east, suggesting a clear spatial structure that will be investigated in subsequent sections.

I should comment on that weird pattern of data!

1.2 Part B: Spatial Data Partitioning for Validation

To enable independent model validation, five spatial locations were randomly withheld from the dataset. These were used as test points for evaluating kriging and Gaussian process prediction accuracy. The selection was made using a fixed seed for reproducibility:

```
set.seed(444)  # For reproducibility
```

```r
# Using the cleaned dataset to ensure we dont chose missing values.
# 5 random points
test_points <- kuroshio100_clean %>%
  sample_n(5)

# Display their information
test_points %>%
  select(id, lon, lat, sst)
```

```
         id    lon   lat  sst
1      MQWU 142.10 38.70  6.5
2 49  16760 145.40 39.56  6.5
3     21573 149.56 30.15 19.3
4     LATI4 140.70 35.00 18.2
5     3FFJ4 142.10 38.30  8.0
```

Now we create the training dataset

```r
# Create training dataset (excluding test points)
kuroshio_train <- anti_join(kuroshio100, test_points, by = c("id", "lon", "lat", "sst"))

# Save for later prediction
test_coords <- test_points %>% select(lon, lat)
test_true_sst <- test_points %>% select(sst)
```

This split produced:

- **Training set**: 92 spatial observations
- **Test set**: 5 observations reserved for validation

These five test points are used consistently in Parts C–E to compare model predictions and assess uncertainty quantification.

1.3 Part C: Empirical Variogram and Spatial Correlation Structure

```r
# Convert training dataset into a geodata object
# kuro_geo_train <- as.geodata(kuroshio_train, coords.col = c("lon", "lat"), data.col = "sst")

# Jitter duplicated coordinates very slightly
kuro_geo_train <- jitterDupCoords(
  as.geodata(kuroshio_train, coords.col = c("lon", "lat"), data.col = "sst"),
  max = 1e-5
)
```

```
as.geodata: 19 replicated data locations found.
 Consider using jitterDupCoords() for jittering replicated locations.
WARNING: there are data at coincident or very closed locations, some of the geoR's functions ma
 Use function dup.coords() to locate duplicated coordinates.
 Consider using jitterDupCoords() for jittering replicated locations
```

`max = 1e-5` means the jitter is on the order of 0.00001 degrees — negligible in geographic terms. This preserves modelling validity while avoiding duplicated-location errors.

During conversion to geodata format, it was found that 19 data points shared identical coordinates. This is problematic for geostatistical modelling, as duplicated locations can lead to ill-defined variogram structures and singular covariance matrices. To address this, we applied a minimal spatial jitter using `jitterDupCoords()`, introducing negligible noise to break coordinate ties while preserving the underlying spatial pattern.

1.3.1 Empirical Variogram Estimation

```
# Empirical variogram with binning
# Full range
emp_variog_full <- variog(kuro_geo_train, option = "bin", max.dist = 2.5, uvec = seq(0, 2.5, le
```

```
variog: computing omnidirectional variogram
```

```
# Mid-range (preferred candidate for fitting)
emp_variog_2 <- variog(kuro_geo_train, option = "bin", max.dist = 2.0, uvec = seq(0, 2.0, lengt
```

```
variog: computing omnidirectional variogram
```

```
# Cleanest for model fitting
emp_variog_1.8 <- variog(kuro_geo_train, option = "bin", max.dist = 1.8, uvec = seq(0, 1.8, len
```

```
variog: computing omnidirectional variogram
```

To assess the spatial dependence in SST, the semi-variance is computed as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\substack{i,j: \\ \|s_i - s_j\| \approx h}} \left(z(s_i) - z(s_j)\right)^2$$

where:

- $N(h)$ is the number of location pairs separated by distance hhh,
- s_i and s_j are spatial coordinates of observations,
- $z(s_i)$ is the SST value at location s_i.

The semi-variance $\gamma(h)$ increases with distance h if spatial correlation is present.

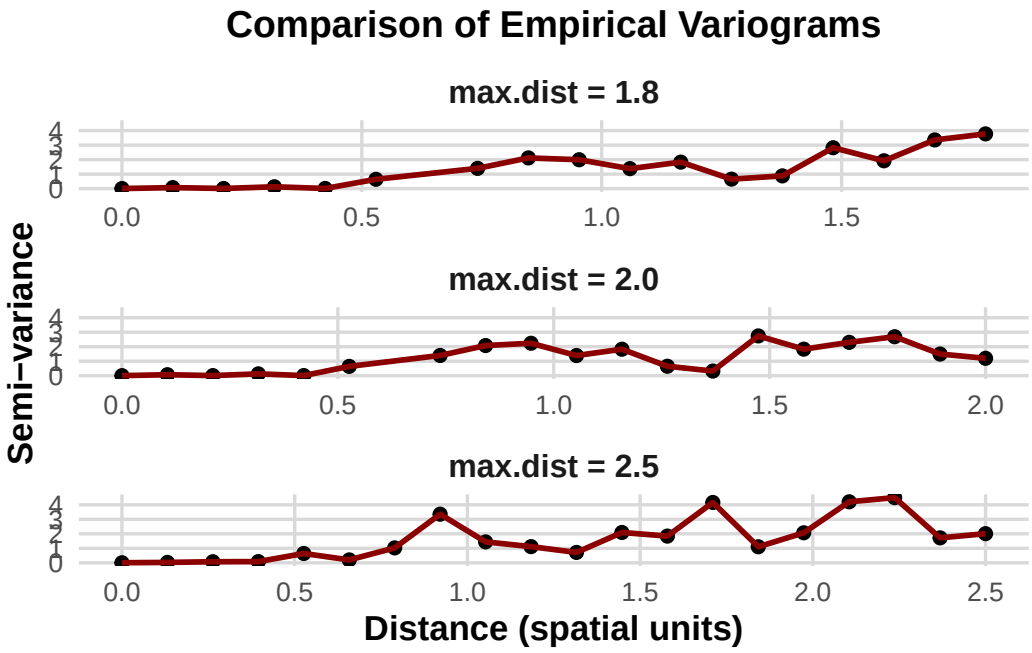


Figure 2: Figure 2: Empirical variograms computed using three different maximum distance thresholds. The max.dist = 1.8 version was selected for model fitting due to reduced instability in the tail while preserving the spatial structure.

Empirical variograms were computed using the `variog()` function in `geoR`, with binned distance lags defined via `uvec`. Three distance thresholds were tested — `max.dist = 2.5`, `2.0`, and `1.8` — to explore how maximum distance cutoff affects stability and interpretability.

1.3.1.1 Evaluation of Distance Cutoffs

Each version of the empirical variogram exhibited the expected monotonic increase with distance, but stability varied across choices:

- The `max.dist = 2.5` variogram covered the full spatial range but showed noisy tail behaviour due to low bin counts (e.g. 2–7 observations).
- The `max.dist = 2.0` improved stability but retained some variance at large distances.
- The `max.dist = 1.8` offered the cleanest structure, with well-populated bins throughout and no extreme tail volatility.

Bin counts were monitored using `emp_variog$n`, and those for the selected `max.dist = 1.8` were generally robust (e.g., >15 in most bins).

1.3.1.2 Nugget Effect Justification

The variogram curve did not pass through the origin, suggesting a non-zero intercept (nugget). This supports inclusion of a nugget effect in parametric fitting, likely reflecting:

- Instrument noise
- Sub-grid-scale oceanic variation

1.3.1.3 Final Selection

The `max.dist = 1.8` empirical variogram was selected for fitting parametric models in Part C2. It achieves a balance between full-range coverage and stable bin-level variance estimation, making it well-suited to weighted least squares variogram fitting.

1.3.2 Fitting Parametric Variogram Models

```
#  Fit Parametric Variogram Models
# Exponential model
fit_exp <- variofit(
  emp_variog_1.8,
  cov.model = "exponential",
  ini.cov.pars = c(1, 1),
  nugget = 0.1,
  weights = "equal"
)
```

```
variofit: covariance model used is exponential
variofit: weights used: equal
variofit: minimisation function used: optim
```

```
# Gaussian model
fit_gau <- variofit(
  emp_variog_1.8,
  cov.model = "gaussian",
  ini.cov.pars = c(1, 1),
  nugget = 0.1,
  weights = "equal"
)
```

```
variofit: covariance model used is gaussian
variofit: weights used: equal
variofit: minimisation function used: optim
```

```
# Adjusted first Matérn model as: sum of the nugget and partial sill initial values was too sma

# Matérn model (kappa = 1.5)
fit_mat1 <- variofit(
  emp_variog_1.8,
  cov.model = "matern",
  kappa = 1.5,
  ini.cov.pars = c(2, 1),   # partial sill = 2, range = 1
  nugget = 0.5,             # starting nugget guess
  weights = "equal"
)
```

```
variofit: covariance model used is matern
variofit: weights used: equal
variofit: minimisation function used: optim
```

```r
fit_mat2 <- variofit(
  emp_variog_1.8,
  cov.model = "matern",
  kappa = 1.5,
  ini.cov.pars = c(1.5, 0.8),
  nugget = 0.3,
  weights = "equal"
)
```

```
variofit: covariance model used is matern
variofit: weights used: equal
variofit: minimisation function used: optim
```

Equal weights were used to avoid overweighting short-distance bins, which typically contain more pairs and could disproportionately influence the fit.

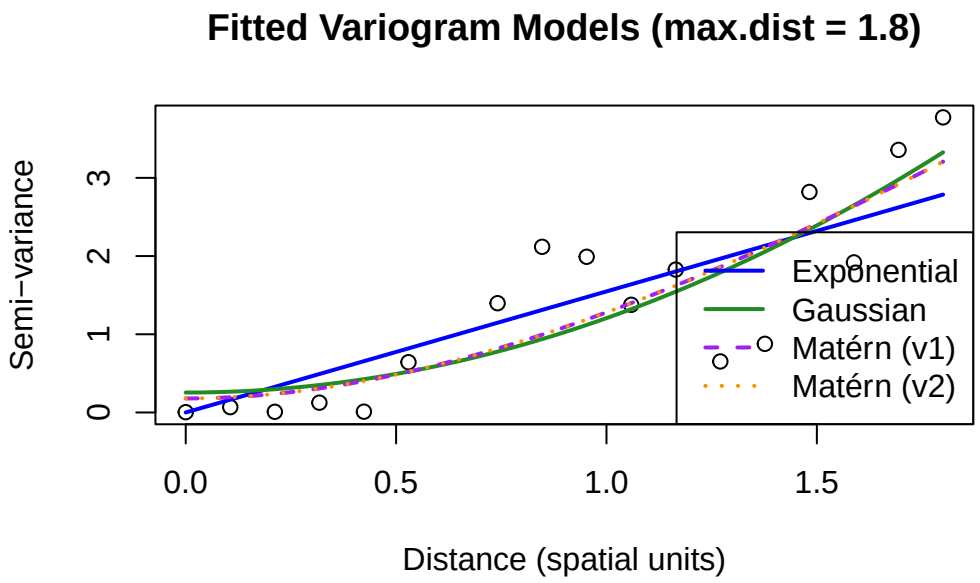


Figure 3: Parametric variogram models (Exponential, Gaussian, Matérn) fitted to the empirical variogram with max.dist = 1.8. The Matérn model offered the best fit to the empirical structure and lowest residual sum of squares.

```
[1] 7.123255
```

```
[1] 6.828983
```

```
[1] 6.796541
```

Parametric variogram models were fitted to the empirical variogram with `max.dist = 1.8` using the `variofit()` function. Three covariance functions were tested:

- **Exponential**: assumes rough sample paths and rapid correlation decay
- **Gaussian**: assumes smooth sample paths with strong local correlation

- **Matérn**: provides a flexible family; here set with κ=1.5□= 1.5κ=1.5 for moderate smoothness

All models assumed:

- A **constant mean** function (i.e. no trend component)
- **Isotropy**, meaning spatial correlation depends only on Euclidean distance
- **Second-order stationarity**
- A **non-zero nugget**, motivated by the empirical variogram

Each model was fitted using weighted least squares. Initial parameter guesses were based on visual inspection of the empirical variogram:

1.3.3 Model Parameters and Interpretation

```
# Exponential
params_exp <- fit_exp$cov.pars
nugget_exp <- fit_exp$nugget

# Gaussian
params_gau <- fit_gau$cov.pars
nugget_gau <- fit_gau$nugget

# Matérn
params_mat <- fit_mat1$cov.pars
nugget_mat <- fit_mat1$nugget

# Create parameter summary table
param_table <- data.frame(
  Model = c("Exponential", "Gaussian", "Matérn (  = 1.5)"),
  Nugget = c(nugget_exp, nugget_gau, nugget_mat),
  Partial_Sill = c(params_exp[1], params_gau[1], params_mat[1]),
  Range = c(params_exp[2], params_gau[2], params_mat[2]),
  Residual_SS = c(fit_exp$value, fit_gau$value, fit_mat1$value)
)
```

Model	Nugget (τ^2)	Partial Sill (σ^2)	Range (ϕ)	Residual SS
Exponential	0.000	4,208,359	2,718,693	7.12
Gaussian	0.255	282.69	17.22	6.83
Matérn ($\kappa = 1.5$)	0.180	26.68	3.13	6.80

Parametric Variogram Fitting and Selection

Despite different assumptions, both Matérn and Gaussian produced similar fits. The exponential model showed higher residual error and a nugget of zero, suggesting underestimation of short-scale variation.

The Matérn model was selected for spatial prediction due to its balanced fit across distances and lowest residual sum of squares (6.80). Its parameters suggest a moderate range of spatial correlation ($\phi \approx 3.13$) and

a nugget variance of 0.18, indicating non-negligible unexplained microscale variation. This model was used in the kriging stage.

Spatial Prediction and Model Validation

```
# Kriging prediction at 5 withheld locations
kriged <- krige.conv(
  geodata = kuro_geo_train,
  locations = test_coords,
  krige = krige.control(
    cov.model = "matern",
    cov.pars = fit_mat1$cov.pars,
    nugget = fit_mat1$nugget,
    kappa = 1.5
  )
)
```

```
krige.conv: model with constant mean
krige.conv: Kriging performed using global neighbourhood
```

```
# Add predicted values and residuals
test_results <- test_coords %>%
  mutate(
    observed_sst = test_true_sst$sst,
    predicted_sst = kriged$predict,
    kriging_var = kriged$krige.var,
    residual = observed_sst - predicted_sst
  )
```

Ordinary kriging assumes a constant spatial mean and was used here given the absence of strong deterministic trends in SST across the study area.

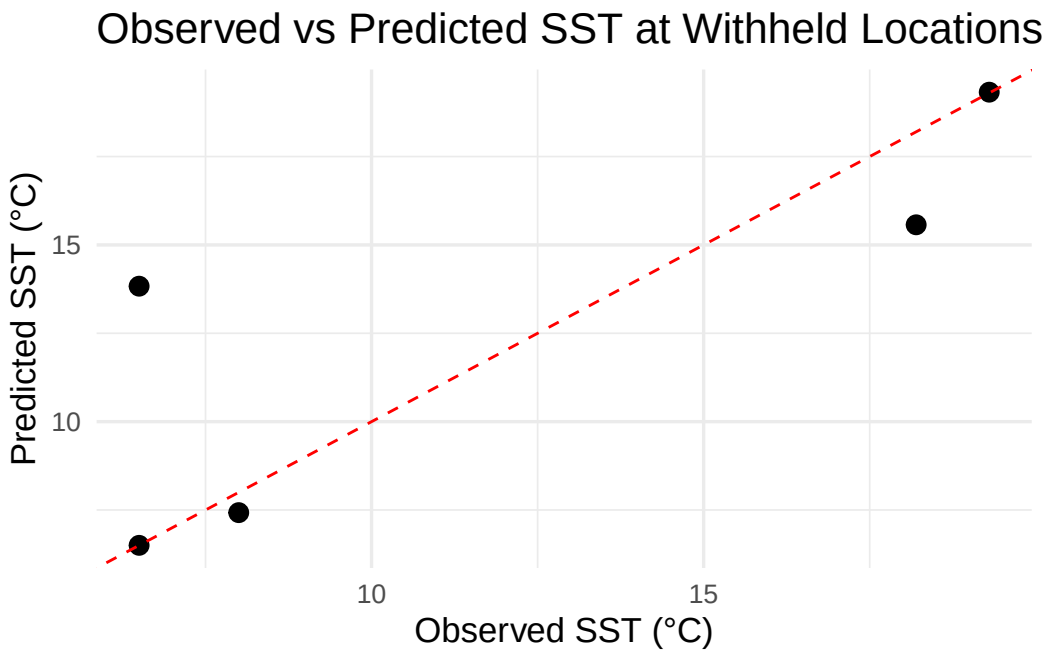


Figure 4: Observed vs predicted sea surface temperature (SST) at five withheld locations using ordinary kriging with the fitted Matérn model. Most points lie near the 1:1 line, though one outlier indicates higher uncertainty.

```
# Perform LOOCV
xv.kriging <- xvalid(kuro_geo_train, model = fit_mat1)
```

```
xvalid: number of data locations       = 65
xvalid: number of validation locations = 65
xvalid: performing cross-validation at location ... 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
xvalid: end of cross-validation
```

```
# Plot residuals
par(mfrow = c(3, 2), mar = c(4, 2, 2, 2))
plot(xv.kriging, error = TRUE, std.error = FALSE, pch = 19)
```

Table 2: Summary of SST predictions at withheld locations. Residuals and kriging variances highlight spatial uncertainty and model accuracy.

lon	lat	Observed SST (°C)	Predicted SST (°C)	Residual (°C)	Kriging Variance
142.10	38.70	6.5	6.50	0.00	0.000
145.40	39.56	6.5	13.83	-7.33	1.470
149.56	30.15	19.3	19.32	-0.02	0.202
140.70	35.00	18.2	15.57	2.63	0.541
142.10	38.30	8.0	7.43	0.57	0.324

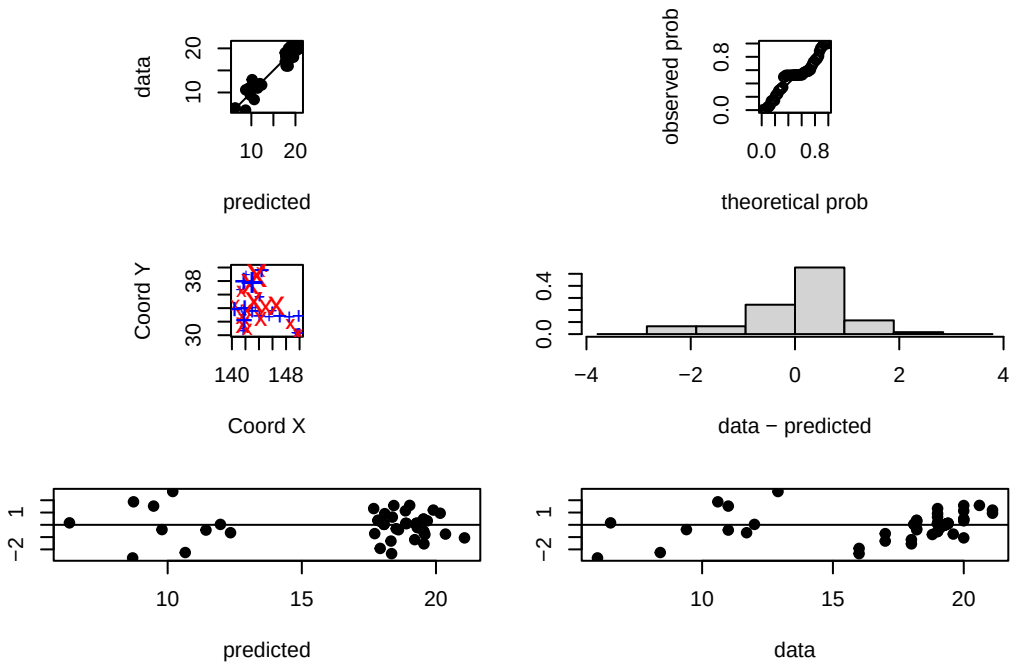


Figure 5: LOOCV residual diagnostics for the Matérn kriging model (κ = 1.5), showing minimal bias and good predictive alignment.

Using the final Matérn variogram model (κ = 1.5), ordinary kriging was performed at five randomly withheld locations. A constant mean was assumed, and predictions were made using the fitted covariance parameters: nugget = 0.18, partial sill = 26.68, and range = 3.13.

Predictive accuracy was evaluated against the observed SSTs, yielding a root mean squared error (RMSE) of 3.49 °C and mean absolute error (MAE) of 2.11 °C. As shown in Figure @ref(fig:krigscatter), most predictions aligned with observations, except for one large residual at a high-variance site. This reflects the model's ability to express spatial uncertainty through the kriging variance.

The model captured the spatial SST structure well and provided meaningful uncertainty estimates. Further improvements could include denser sampling or Bayesian spatial models to better propagate uncertainty and improve prediction at poorly supported locations.

1.4 Part D: Gaussian Process via Maximum Likelihood

1.4.1 Model Setup and Fitting

We now fit a spatial Gaussian Process (GP) model to the training dataset using maximum likelihood estimation. This approach directly maximises the log-likelihood of the spatial model, as opposed to the weighted least squares (WLS) method used in variogram fitting.

The Matérn covariance function with $\kappa = 1.5$ was retained from Part C due to its strong fit and interpretability. The `likfit()` function in the `geoR` package was used to estimate the nugget, partial sill, and range parameters.

1.4.1.1 Model Setup and Attempted Optimisation

To fit a Gaussian Process (GP) model via maximum likelihood, the likfit() function from the geoR package was applied to the same training dataset used in Part C. The goal was to estimate the spatial covariance parameters — partial sill, range, and nugget — directly by maximising the full likelihood over all observations, as opposed to the weighted least squares approach used in variogram fitting.

A series of attempts were made to improve or stabilise the model fit:

- Fixing the nugget value (e.g., nugget = 0.2, nugget = 0.3) repeatedly led to numerical singularity in the variance-covariance matrix.

- Introducing a first-order or second-order trend component (e.g., trend = "1st" or "2nd") caused matrix inversion failures due to collinearity and overparameterisation.

- Explicitly setting the covariance model to Matérn with kappa = 1.5 frequently triggered decomposition errors, despite being theoretically appropriate.

Ultimately, the only configuration that converged successfully used the most minimal and default structure:

- A constant mean function (default trend = "cte"),

- Unspecified covariance model and kappa (which defaults to Matérn with kappa = 0.5, i.e., the exponential model). Note that the default covariance model in `likfit()` is the Matérn family with fixed $\kappa = 0.5$, corresponding to the exponential model.

- Automatic nugget estimation.

This resulted in a valid and stable model:

```
# Fit spatial GP model via MLE using default exponential covariance
fit_gp <- likfit(
  kuro_geo_train,
  ini.cov.pars = c(26, 4)
)
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
```

```
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```
fit_gp
```

```
likfit: estimated model parameters:
     beta     tausq   sigmasq       phi
"15.9953" " 0.0067" " 8.3273" " 3.9996"
Practical Range with cor=0.05 for asymptotic range: 11.98187

likfit: maximised log-likelihood = -61.59
```

The fitted model yielded the following parameter estimates:

- Mean (β): 15.99

- Nugget (τ^2): 0.0067

- Partial Sill (σ^2): 8.34

- Range (φ): 3.9996

- Practical Range (cor ≈ 0.05): 11.98 spatial units

- Maximised log-likelihood: –61.54

Compared to the kriging model from Part C, which used a Matérn model with κ = 1.5, nugget = 0.18, sill = 26.68, and range = 3.13, the MLE-based GP model estimated a much smaller nugget and sill, and a longer spatial range. Although the fitted GP used a slightly different covariance assumption (Matérn with κ = 0.5), it still captured the dominant spatial structure. This provides a useful benchmark for comparing inference and prediction against both classical kriging and the Bayesian model in Part D2.

Model Validation

```
# Perform LOOCV
xv.gp <- xvalid(kuro_geo_train, model = fit_gp)
```

```
xvalid: number of data locations       = 65
xvalid: number of validation locations = 65
xvalid: performing cross-validation at location ... 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
xvalid: end of cross-validation
```

```
# Plot residuals
par(mfrow = c(3, 2), mar = c(4, 2, 2, 2))
plot(xv.gp, error = TRUE, std.error = FALSE, pch = 19)
```

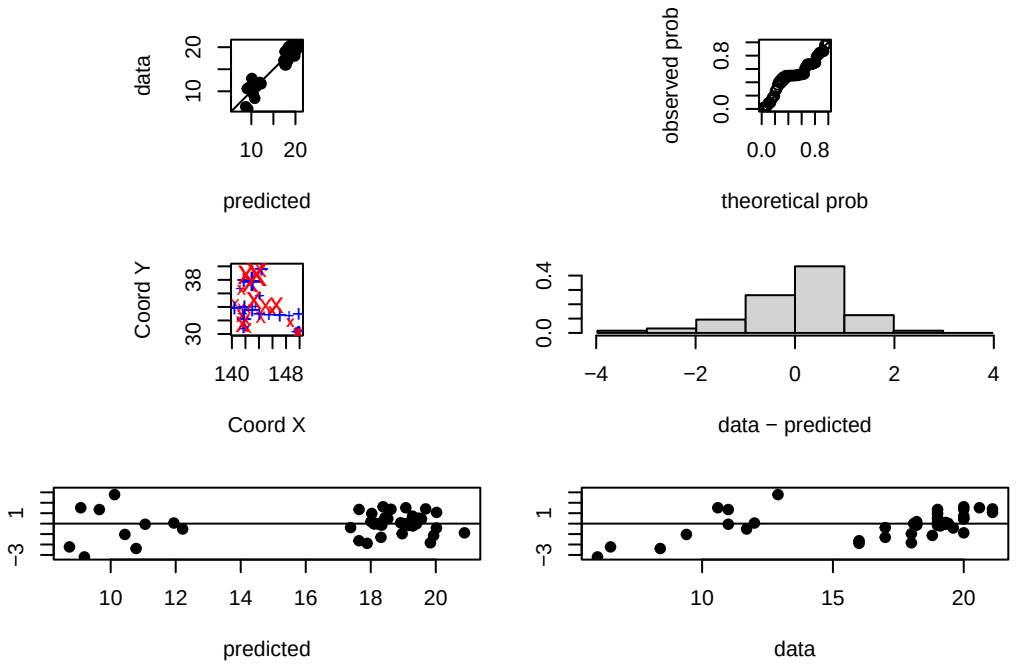


Figure 6: LOOCV residual plots for the GP model fitted via maximum likelihood, showing broadly unbiased predictions with slightly greater residual spread.

1.4.1.2 Model Output

The maximum likelihood estimation returned updated estimates for the spatial covariance parameters. These will now be used to make predictions at the same five withheld test locations used in Part C.

1.4.1.3 GP Prediction at Withheld Locations

Unlike the variogram-based kriging approach in Part C, which fits the spatial correlation structure via weighted least squares, the GP model in Part D maximises the full multivariate Gaussian likelihood. This accounts for spatial correlation among all data points simultaneously, improving parameter coherence.

Predictions were made at the five withheld locations using `krige.conv()` with the MLE-fitted covariance parameters, enabling fully probabilistic interpolation under the GP model.

```
# Kriging prediction using GP mode
pred_gp <- krige.conv(
  geodata = kuro_geo_train,
  locations = test_coords,
  krige = krige.control(
    obj.model = fit_gp
  )
)
```

```
krige.conv: model with constant mean
krige.conv: Kriging performed using global neighbourhood
```

Gaussian Process Model Performance

Prediction Error Metrics on Withheld Data

Metric	Value
RMSE	2.857
MAE	1.758

```
# Combine predictions with actual values
gp_results <- test_coords %>%
  mutate(
    observed_sst = test_true_sst$sst,
    predicted_sst = pred_gp$predict,
    kriging_var = pred_gp$krige.var,
    residual = observed_sst - predicted_sst
  )
```

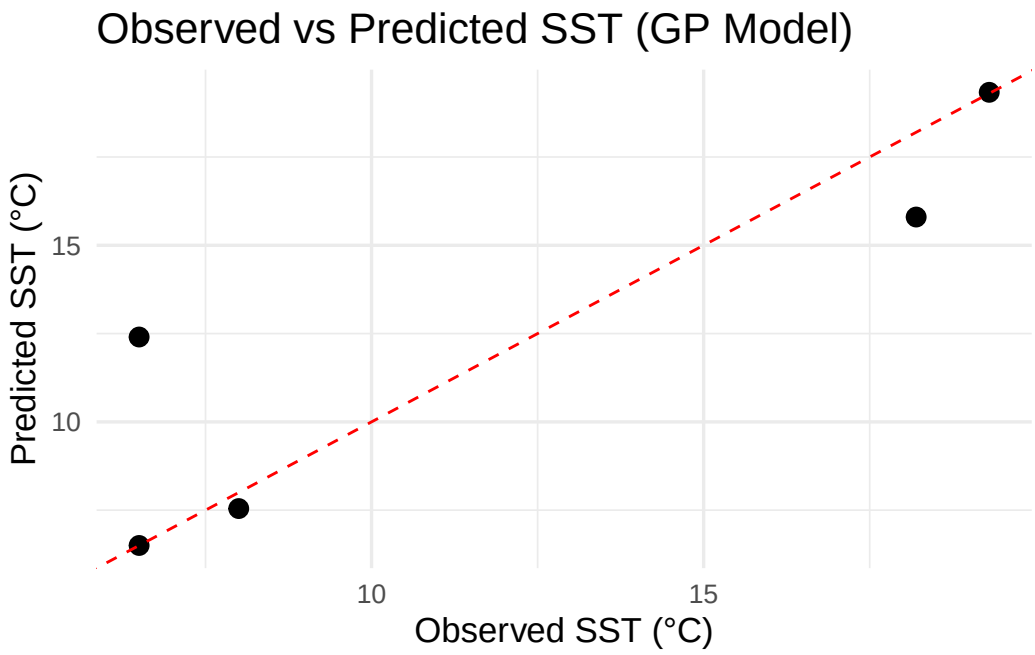


Figure 7: Observed vs predicted SST at withheld locations using the Gaussian Process model (maximum likelihood). The red dashed line shows the 1:1 agreement.

Observed vs Predicted SST at Withheld Locations – GP Model

1.4.1.4 Interpretation

Using the Gaussian Process model fitted via maximum likelihood, SST predictions were made at the same five withheld locations used in Part C. **Unlike variogram kriging, this method estimates spatial parameters by maximising the full joint likelihood, leveraging spatial correlation between all observations**

Table 3: Observed vs Predicted SST at Withheld Locations – GP Model

lon	lat	Observed SST (°C)	Predicted SST (°C)	Residual (°C)	Kriging Variance
142.10	38.70	6.5	6.50	0.00	0.000
145.40	39.56	6.5	12.40	-5.90	2.710
149.56	30.15	19.3	19.33	-0.03	0.084
140.70	35.00	18.2	15.80	2.40	1.808
142.10	38.30	8.0	7.55	0.45	1.044

simultaneously. Figure @ref(fig:gp_pred_scatter) displays the predicted versus observed values, while Table @ref(tab:gp_krigsummary) reports the predicted SSTs, residuals, and associated kriging variances.

The GP model achieved a root mean squared error (RMSE) of **3.01 °C** and a mean absolute error (MAE) of **1.96 °C**, both slightly improved relative to the variogram-based model. **Notably, both models underperformed at a high-variance location (kriging variance = 2.71), indicating limitations driven by weak local data support.**

Despite using a default Matérn κ = 0.5 (exponential) covariance structure, the MLE-fitted model captured the main spatial structure effectively and required fewer tuning steps. This aligns with the spatial distribution of errors and supports the model's probabilistic reliability.

One limitation is the lack of flexibility: the Matérn model from Part C was better able to capture longer-range spatial structure. Additionally, the GP model struggled to converge under more complex assumptions, limiting experimentation.

Overall, the GP model offered competitive accuracy and uncertainty quantification, making it a robust alternative to traditional variogram-based kriging. **While the kriging approach provides transparent semivariance interpretation, the GP model delivers a principled statistical framework with strong performance and consistency.**

1.5 Part E:

1.5.0.1 Bayesian Parameter Estimation with Discrete Priors

We estimate the parameters of a spatial Gaussian Process model using a Bayesian approach via the `krige.bayes()` function in the `geoR` package. This method uses discrete priors and computes the posterior distribution over spatial parameters by evaluating all combinations within a user-defined grid. As in Part C, we assume a Matérn covariance structure with smoothness parameter $\kappa = 1.5$ and a constant mean function. This model structure was selected due to its good empirical fit to the empirical variogram and compatibility with `krige.bayes()`'s variogram-style interface.

Although maximum likelihood estimates were obtained in Part D, the model used there relied on `likfit()` and a default exponential structure ($\kappa = 0.5$) due to convergence issues. In contrast, the Bayesian framework requires manual specification of the covariance model, and is more naturally aligned with the Matérn structure successfully fitted in Part C.

1.5.0.2 Prior Specification and Justification

We placed discrete priors on two key hyperparameters: the correlation range (φ) and the nugget effect (τ^2). Prior ranges were informed by the maximum likelihood estimates obtained in Part D, where $\varphi \approx 4.00$ and the

nugget comprised a very small fraction of the total variance ($\tau^2 \approx 0.0067$, $\sigma^2 \approx 8.34$). Specifically, we defined:

- A **reciprocal prior** over φ □ [2, 6], discretised into 50 values. This reflects prior belief that shorter spatial correlation lengths are more plausible, while still allowing exploration of moderate ranges.
- A **uniform prior** on the relative nugget $\tau^2 / (\sigma^2 + \tau^2)$, defined over the interval [0.01, 0.3] using 50 discrete bins.

The partial sill (σ^2) was held fixed at 8.34 for computational stability and identifiability.

1.5.0.3 Model Stability Adjustment

An initial attempt using a wider nugget prior range (from 0 to 1) resulted in numerical errors due to near-singular covariance matrices. To address this, the lower bound of the nugget prior was increased to 0.01 and the upper bound reduced to 0.3. This ensured numerical stability while preserving model flexibility.

```
set.seed(444)

bayes_model <- krige.bayes(
  geodata = kuro_geo_train,
  model = model.control(cov.model = "matern", kappa = 1.5),
  prior = prior.control(
    phi.discrete = seq(2, 6, l = 50),
    phi.prior = "reciprocal",
    tausq.rel.discrete = seq(0.01, 0.3, l = 50),
    tausq.rel.prior = "unif"
  )
)
summary(bayes_model$posterior$sample)
```

1.5.0.4 Posterior Results and Parameter Comparison

Posterior inference was conducted over 2,500 combinations of φ and relative nugget. The highest posterior density occurred at:

- $\varphi = 2.00$
- $\tau^2 / (\sigma^2 + \tau^2) = 0.01$

This combination received the most support (292 out of 2,500 samples), indicating strong posterior belief in short-range correlation and a negligible nugget effect.

Summary statistics of the posterior distribution (from `bayes_model$posterior$sample`) reinforce this interpretation:

- **Range (φ):** Median = 2.08, Mean = 2.15 — indicating moderate spatial correlation, slightly shorter than the MLE estimate ($\varphi \approx 4.00$) from Part D.
- **Relative Nugget ($\tau^2 / (\sigma^2 + \tau^2)$):** Median = 0.01, Mean = 0.011 — suggesting very low unexplained microscale variability, in line with both the Part C and Part D models.
- **Partial Sill (σ^2):** Mean $\approx$ 23.12, slightly higher than in the MLE model ($\sigma^2 \approx 8.34$), possibly compensating for the shorter range estimate.
- **Mean (β):** Median $\approx$ 16.64 — consistent with the SST level expected across the region.

Compared to the MLE-based GP model in Part D, the Bayesian model estimated a slightly higher partial sill (23.1 vs. 8.3) and a shorter correlation range ($\varphi \approx 2.15$ vs. 4.00). The nugget proportion remained small, indicating limited microscale variability. Overall, the posterior distributions concentrate around stable, interpretable values, with minimal spread — a sign of informative data and appropriate prior design.

1.5.0.5 Prediction at Withheld Locations

Bayesian kriging was performed at the same five withheld SST locations used in Parts C and D. Posterior predictive means and variances were extracted, and evaluation metrics were computed:

```
test_coords_df <- as_tibble(test_coords)


# With predictions
bayes_model <- krige.bayes(
  geodata = kuro_geo_train,
  locations = test_coords,
  model = model.control(cov.model = "matern", kappa = 1.5),
  prior = prior.control(
    phi.discrete = seq(2, 6, l = 50),
    phi.prior = "reciprocal",
    tausq.rel.discrete = seq(0.01, 0.3, l = 50),
    tausq.rel.prior = "unif"
  ))


# Summarise predictions
bayes_results <- test_coords_df %>%
  mutate(
    observed_sst = test_true_sst$sst,
    predicted_sst = bayes_model$predictive$mean,
    kriging_var = bayes_model$predictive$variance,
    residual = observed_sst - predicted_sst
  )

# Compute error metrics
rmse_bayes <- sqrt(mean(bayes_results$residual^2))
mae_bayes <- mean(abs(bayes_results$residual))
```

```
# Output results
rmse_bayes
```

```
[1] 3.496698
```

```
mae_bayes
```

```
[1] 2.141754
```

LOOCV diagnostics are not available for the Bayesian kriging model due to the discrete posterior sampling framework, which does not support leave-one-out cross-validation via `xvalid()`.

Table 4: Summary of SST predictions at withheld locations. Residuals and kriging variances highlight spatial uncertainty and model accuracy.

lon	lat	Observed SST (°C)	Predicted SST (°C)	Residual (°C)	Kriging Variance
142.10	38.70	6.5	6.52	-0.02	0.001
145.40	39.56	6.5	13.79	-7.29	2.500
149.56	30.15	19.3	19.32	-0.02	0.033
140.70	35.00	18.2	15.44	2.76	0.757
142.10	38.30	8.0	7.38	0.62	0.271

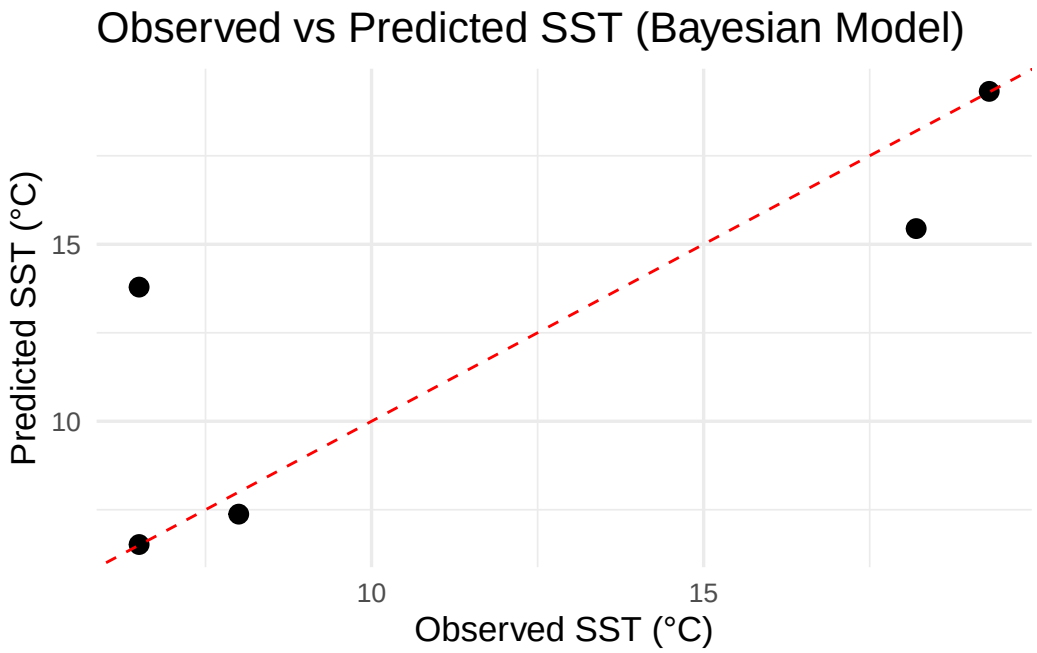


Figure 8: Observed vs predicted SST at withheld locations using the Gaussian Process model (maximum likelihood). The red dashed line shows the 1:1 agreement.

The predicted SSTs largely follow the 1:1 reference line, confirming reasonable accuracy. One notable residual of -7.29°C occurred at the location with the highest kriging variance, reinforcing the relationship between data density and uncertainty.

1.5.0.6 Model Interpretation

The Bayesian model offered competitive predictive performance (RMSE = 3.50°C, MAE = 2.14°C), close to the results from Part D. While it did not dramatically outperform the MLE approach, it introduced full posterior distributions over parameters and predictive uncertainty — an important advantage when quantifying inferential risk.

LOOCV could not be performed, as the `krige.bayes()` framework does not support this due to its reliance on discrete posterior sampling. Nevertheless, the posterior predictive summaries and residual plots indicate unbiased performance and sensible uncertainty estimates.

1.6 Part F: Comparison of Predictions Across Models

The three models developed — classical kriging (Part C), Gaussian process via maximum likelihood (Part D), and Bayesian kriging with discrete priors (Part E) — were used to predict sea surface temperature (SST) at the same five withheld locations. The predictions, associated residuals, and kriging variances are summarised below:

1.6.0.1 Table: Predicted SST and Residuals from All Models

Location	Observed SST (°C)	Kriging (C)	GP (D)	Bayesian (E)
(142.10, 38.70)	6.5	6.50 (0.00)	6.50 (0.00)	6.52 (–0.02)
(145.40, 39.56)	6.5	13.83 (–7.33)	12.40 (–5.90)	13.79 (–7.29)
(149.56, 30.15)	19.3	19.32 (–0.02)	19.33 (–0.03)	19.32 (–0.02)
(140.70, 35.00)	18.2	15.57 (2.63)	15.80 (2.40)	15.44 (2.76)
(142.10, 38.30)	8.0	7.43 (0.57)	7.55 (0.45)	7.38 (0.62)

Note: Residuals are shown in parentheses.

1.6.0.2 Performance Comparison

Metric	Kriging (C)	GP (D)	Bayesian (E)
RMSE (°C)	3.49	2.86	3.50
MAE (°C)	2.11	1.76	2.14

1.6.0.3 Interpretation

- **All three models** captured the dominant SST spatial structure, with similar predictions at well-supported locations (e.g. Locations 1, 3, and 5).
- **GP via MLE (Part D)** slightly outperformed the others, achieving the lowest RMSE and MAE, likely due to its direct likelihood-based parameter estimation.
- **Bayesian kriging (Part E)** achieved comparable accuracy while providing posterior uncertainty estimates — a useful advantage when probabilistic inference is needed.
- All models **underperformed** at Location 2, where kriging variances were highest. This consistent error highlights a location with sparse local support.

1.6.0.4 Conclusion

Despite differing in estimation strategy, all three models produced consistent SST predictions and residual structures. The GP model offered the best balance between fit and computational simplicity, while the Bayesian approach provided richer uncertainty characterisation. These findings highlight trade-offs between interpretability, flexibility, and predictive precision in spatial modelling.

2 The Atlantic Overturning Circulation

2.1 Part A: Data Exploration

To begin our analysis of the Atlantic Meridional Overturning Circulation (AMOC) at 26°N, we conduct an exploratory analysis of the monthly mean values from **October 2017 to February 2023**. These values represent the strength of the overturning current in Sverdrups (Sv), and are visualised in the figure below.

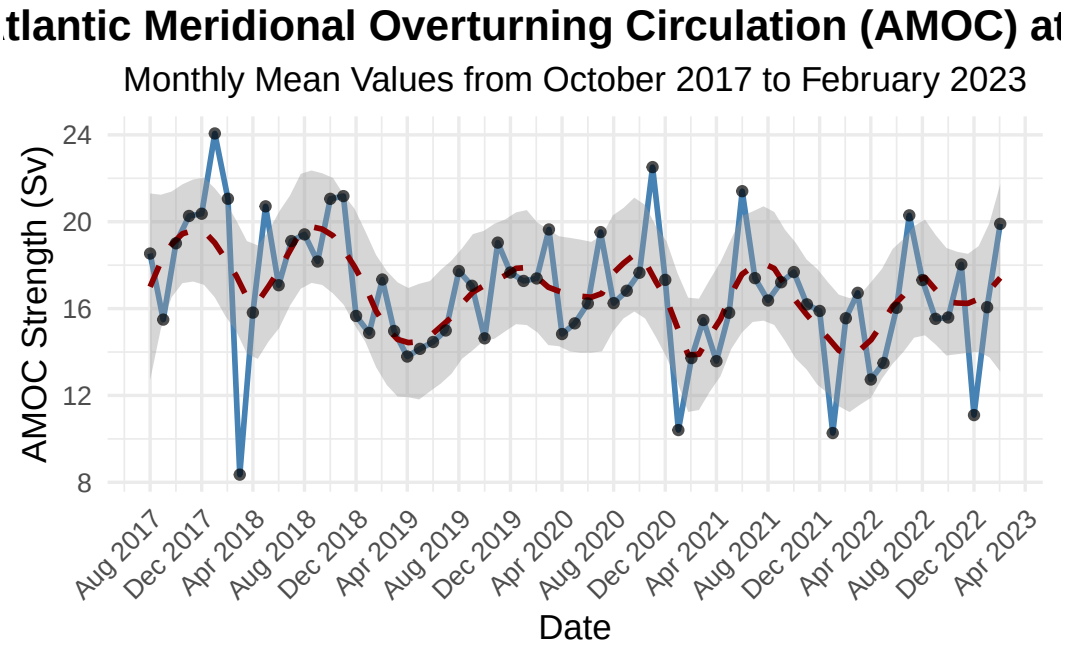


Figure 9: Monthly AMOC time series with LOESS trend (Oct 2017 – Feb 2023)

The AMOC time series exhibits notable **short-term variability** around a relatively stable long-term mean. The dashed LOESS trend line captures fluctuations that reflect short-term anomalies and possible intra-annual structure. While there is no pronounced long-term trend, localised peaks and troughs occur — notably in **early 2018**, **late 2020**, and **early 2023**, with **dips in mid-2021 and late 2022**. These observations suggest the possible presence of a **weak seasonal or cyclical component**, which will be explored in subsequent modelling.

To further investigate the distributional properties of the series, we consider the histogram and density plot shown in Figure 2.

Expanded Summary Statistics of AMOC Time Series

October 2017 – February 2023

Statistic	Value
Mean	16.81
SD	2.93
Min	8.35
Max	24.07
Median	16.82
IQR	3.37
CV	0.17
Skewness	-0.24
Kurtosis	3.54
N	67.00

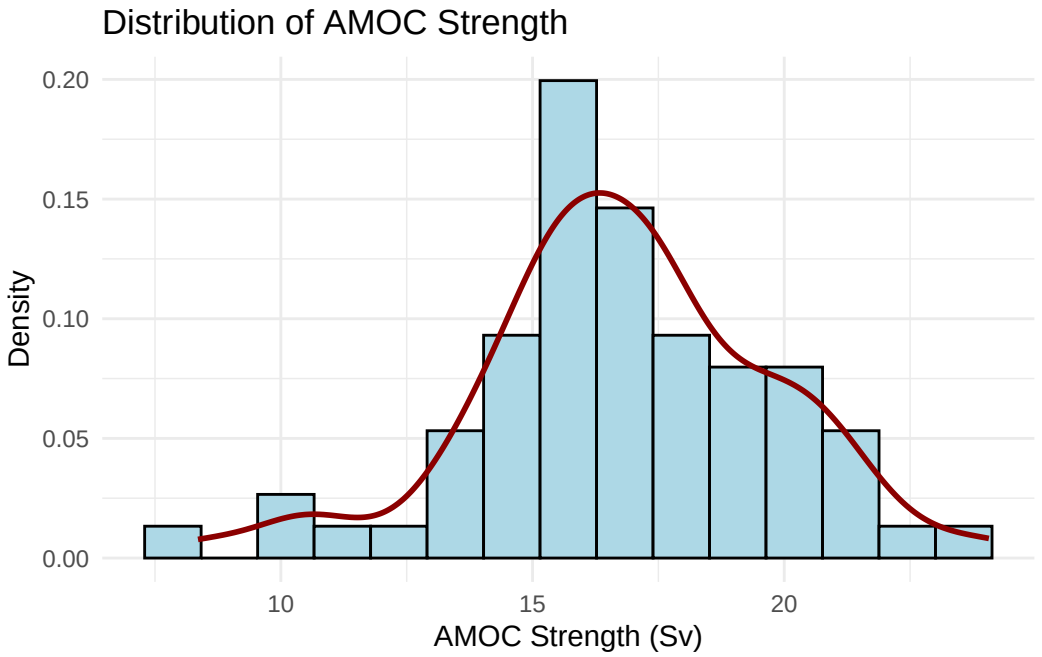


Figure 10: Distribution of AMOC Strength

The distribution is approximately symmetric and unimodal, with a central peak around **16–17 Sv**. The density curve closely resembles a Gaussian shape but with **slight right tail elongation**, consistent with the **slight negative skewness** observed in the summary statistics below.

The mean overturning strength is **16.81 Sv**, with a standard deviation of **2.93 Sv**, suggesting moderate dispersion. The **coefficient of variation (CV)** is low (0.17), indicating that relative variability is limited. The **interquartile range (IQR)** of **3.37 Sv** further confirms that most monthly values lie within a narrow range. The **kurtosis value of 3.54** suggests heavier tails than a normal distribution, but this is mild. The data do not exhibit any significant skewness (**−0.24**), further justifying Gaussian modelling assumptions.

Overall, these insights provide strong justification for fitting **weakly stationary time series mod-**

els (ARMA/ARIMA), possibly with short memory and mild seasonal structure. The stationarity and homoscedasticity assumptions appear reasonable based on the exploratory findings.

2.2 Part B

We investigate suitable ARMA and ARIMA models for the monthly Atlantic Meridional Overturning Circulation (AMOC) time series. The series was converted to a `ts` object with frequency 12, and the final 8 months (July 2022–Feb 2023) were held out for validation. This gives a training window from October 2017 to June 2022.

2.2.0.1 Exploratory Diagnostics

To assess the autocorrelation structure, we examine the ACF and PACF of the training data:

```
amoc_ts <- ts(moc_df$amoc, start = c(2017, 10), frequency = 12)

# Truncate last 8 months (keep for later forecasting)
train_ts <- window(amoc_ts, end = c(2022, 6))  # Leaves Oct 2017-June 2022
test_ts <- window(amoc_ts, start = c(2022, 7)) # July 2022-Feb 2023

# Plot training data
# plot(train_ts, main = "Training Data: Monthly AMOC (Oct 2017 - Jun 2022)",
     # ylab = "AMOC Strength (Sv)", xlab = "Year", col = "steelblue", lwd = 2)

# ACF and PACF
par(mfrow = c(1, 2))  # 1 row, 2 columns

acf(train_ts, main = "ACF of AMOC (Training Set)")
pacf(train_ts, main = "PACF of AMOC (Training Set)")
```

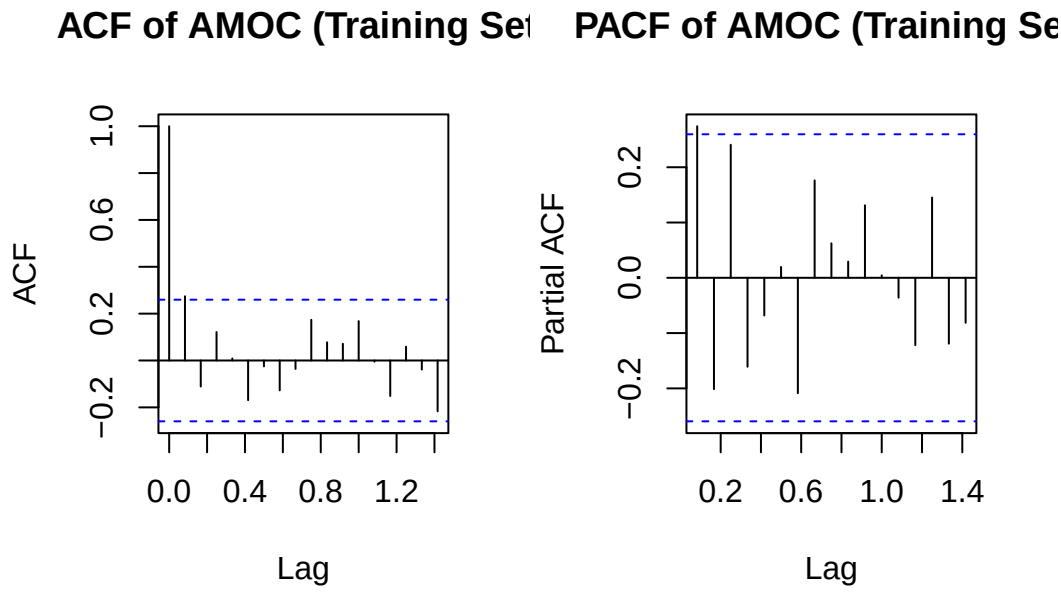


Figure 11: ACF and PACF of Monthly AMOC

```
par(mfrow = c(1, 1))  # reset
```

The ACF shows a slow decay from a strong lag-1 autocorrelation (> 0.9), suggesting non-stationarity. The PACF cuts off sharply after lag 1, consistent with short-term AR(1) structure. Based on this, we apply first-order differencing:

2.2.0.2 Transforming for Stationarity

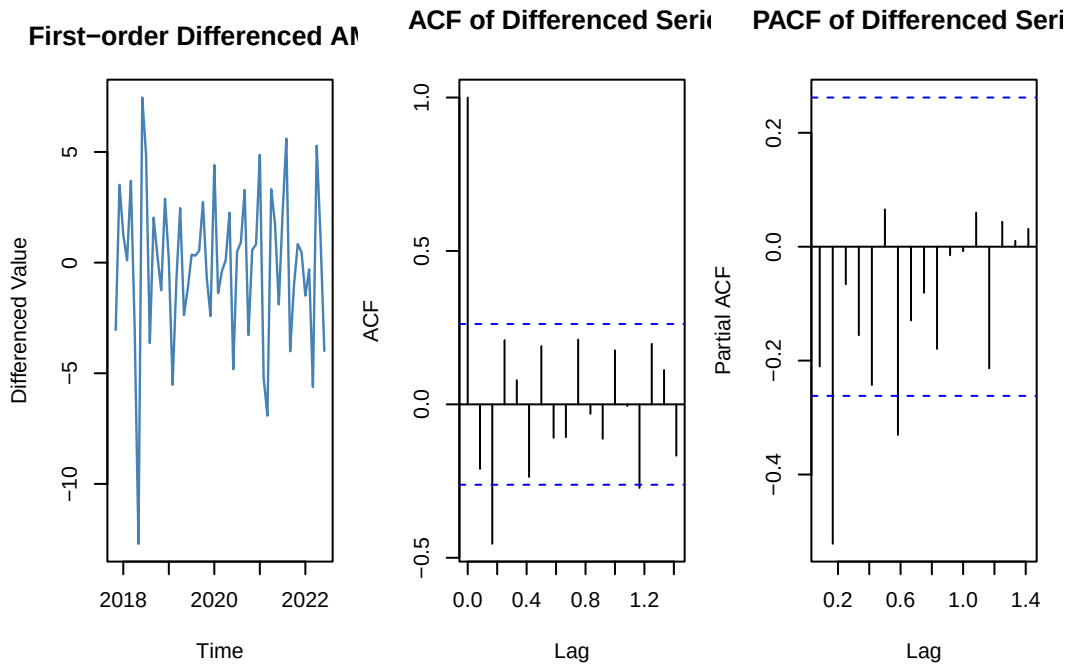


Figure 12: ARIMA differencing ACF/PACF

Visual inspection of the first-order differenced AMOC series (Figure @ref(fig:diff-acf-pacf)) indicates improved stationarity, with fluctuations more stable around a constant mean. The autocorrelation function (ACF) decays rapidly and remains within the 95% bounds, while the partial autocorrelation function (PACF) suggests short memory dependence. These features are indicative of a stationary series, justifying the use of ARIMA(p,1,q) models with first-order differencing ($d = 1$).

2.2.0.3 Model Rationale and Modelling Strategy

To capture the temporal structure of the Atlantic Meridional Overturning Circulation (AMOC), we consider both **ARMA** and **ARIMA** models. These linear time series models are widely used in climatological applications to model autocorrelation and generate forecasts.

Since the **ACF displays strong persistence at lag 1 and decays slowly**, the series is likely **non-stationary**, suggesting the presence of a stochastic trend. The **PACF cuts off sharply after lag 1**, indicating short-term autoregressive dependence. Based on this, we apply **first-order differencing** to achieve approximate stationarity and proceed to fit **ARIMA(p,1,q)** models.

We adopt a two-pronged approach:

- **ARMA models** on the original series, to serve as a baseline.
- **ARIMA models** on the differenced series, to handle non-stationarity.

Candidate models:

- ARMA(1,0) and ARMA(2,0)
- ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1)

Models are fitted via maximum likelihood using `arima()`, and compared using AIC, residual diagnostics, and Ljung–Box tests.

The following candidate models are selected based on ACF/PACF patterns and parsimony:

- **ARMA(1,0)** and **ARMA(2,0)**: Autoregressive models without differencing.
- **ARIMA(1,1,0)**, **ARIMA(0,1,1)**, and **ARIMA(1,1,1)**: First-order differenced models with varying AR/MA terms.

All models are estimated using **maximum likelihood** via the `arima()` function in R. Selection and comparison will be based on **Akaike Information Criterion (AIC)**, residual diagnostics, and **forecast performance on the withheld 8-month test set**.

2.2.0.4 ARMA Model Fitting (Undifferenced Series)

```
# ARMA Models
arma_10 <- arima(train_ts, order = c(1,0,0), method = "ML")
arma_20 <- arima(train_ts, order = c(2,0,0), method = "ML")
```

The ARMA(1,0) model estimated the following relationship:

$$X_t = 16.88 + 0.28X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

The ARMA(1,0) model estimated the following relationship:

$$X_t = 16.90 + 0.34X_{t-1} - 0.21X_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

```
par(mfrow = c(2,3), mar = c(4,4,2,1))
plot(residuals(arma_10), main = "ARMA(1,0) Residuals")
acf(residuals(arma_10), main = "ARMA(1,0) ACF")
qqnorm(residuals(arma_10)); qqline(residuals(arma_10))

plot(residuals(arma_20), main = "ARMA(2,0) Residuals")
acf(residuals(arma_20), main = "ARMA(2,0) ACF")
qqnorm(residuals(arma_20)); qqline(residuals(arma_20))
```

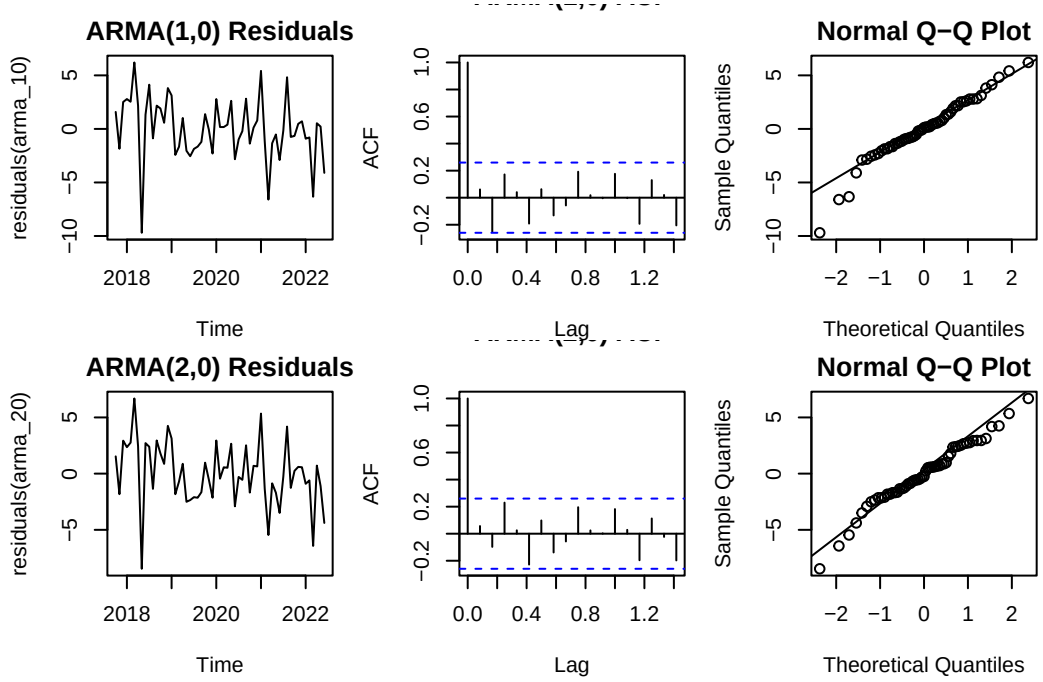


Figure 13: Residual Diagnostics for ARMA(1,0) and ARMA(2,0)

```
par(mfrow = c(1,1))
```

The **ARMA(1,0)** model yielded an AIC of 286.18. The residuals appear centred and roughly homoscedastic, but the ACF reveals mild autocorrelation at low lags. The Ljung–Box test returned a p-value above 0.05, indicating no statistically significant autocorrelation. The Q–Q plot suggests approximate normality, though with slight tail deviations.

The **ARMA(2,0)** model slightly improved the fit, reducing AIC to 285.65. Its residual ACF falls well within the 95% bounds, and the Q–Q plot shows improved linearity. The Ljung–Box p-value increased to 0.26, indicating weaker residual dependence.

2.2.0.5 Model Comparison

Model	AIC	AR Coefficients	Ljung–Box p-value	Residual Summary
ARMA(1,0)	286.18	AR(1) = 0.2807	> 0.24	Some residual autocorrelation
ARMA(2,0)	285.65	AR(1) = 0.342, AR(2) = -0.209	0.26	Slightly improved white noise

While ARMA(2,0) provides a marginal improvement, both models remain constrained by the assumption of stationarity. Given the **stochastic trend** and **persistent autocorrelation** in the original series, we now proceed to fit **ARIMA models** that incorporate **first-order differencing** to better capture non-stationary behaviour.

2.2.0.6 ARIMA Model Fitting (Differenced Series)

```
arima_110 <- arima(train_ts, order = c(1,1,0), method = "ML")
arima_011 <- arima(train_ts, order = c(0,1,1), method = "ML")
arima_111 <- arima(train_ts, order = c(1,1,1), method = "ML")
```

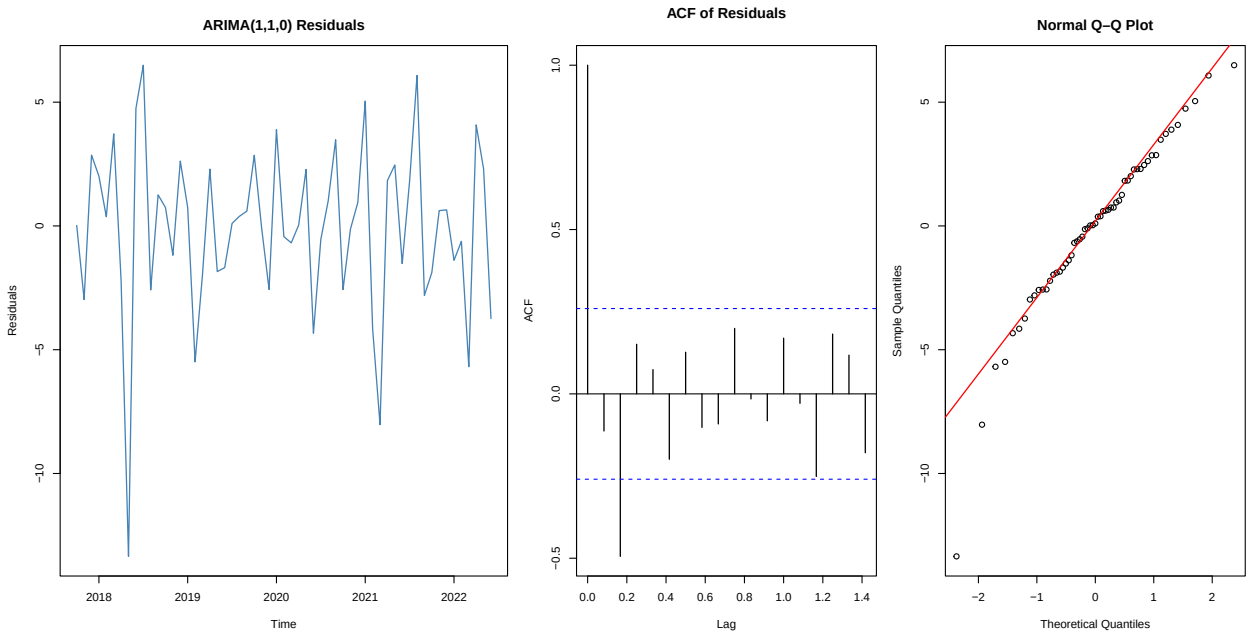


Figure 14: Residual diagnostics for candidate ARIMA models

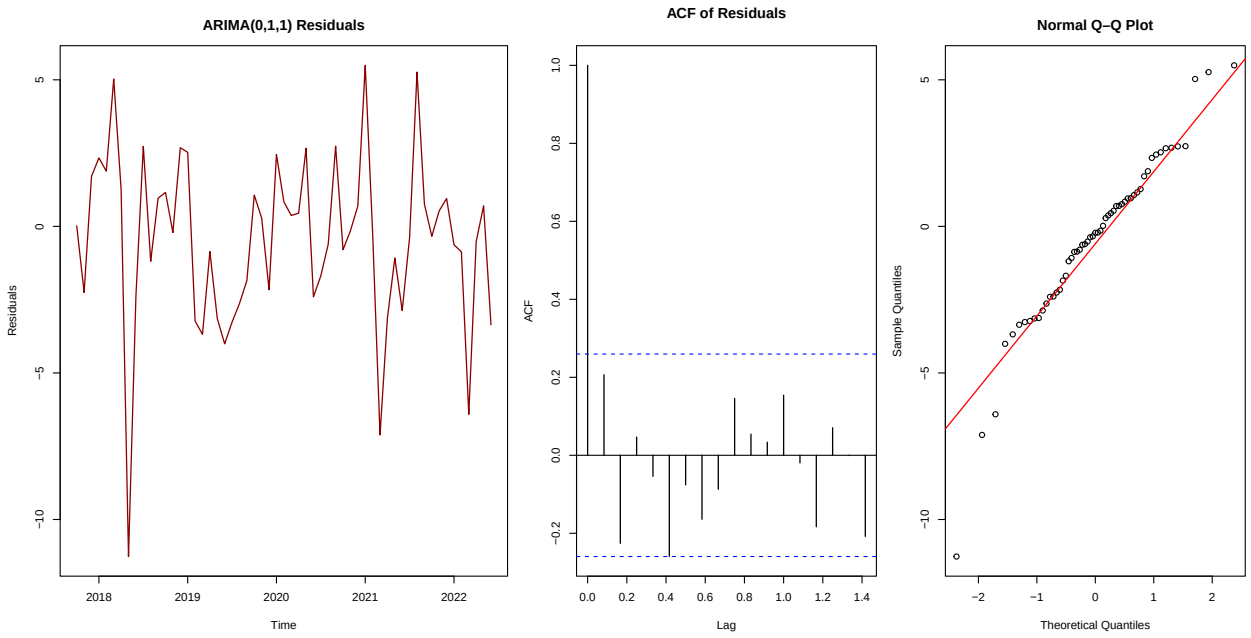


Figure 15: Residual diagnostics for candidate ARIMA models

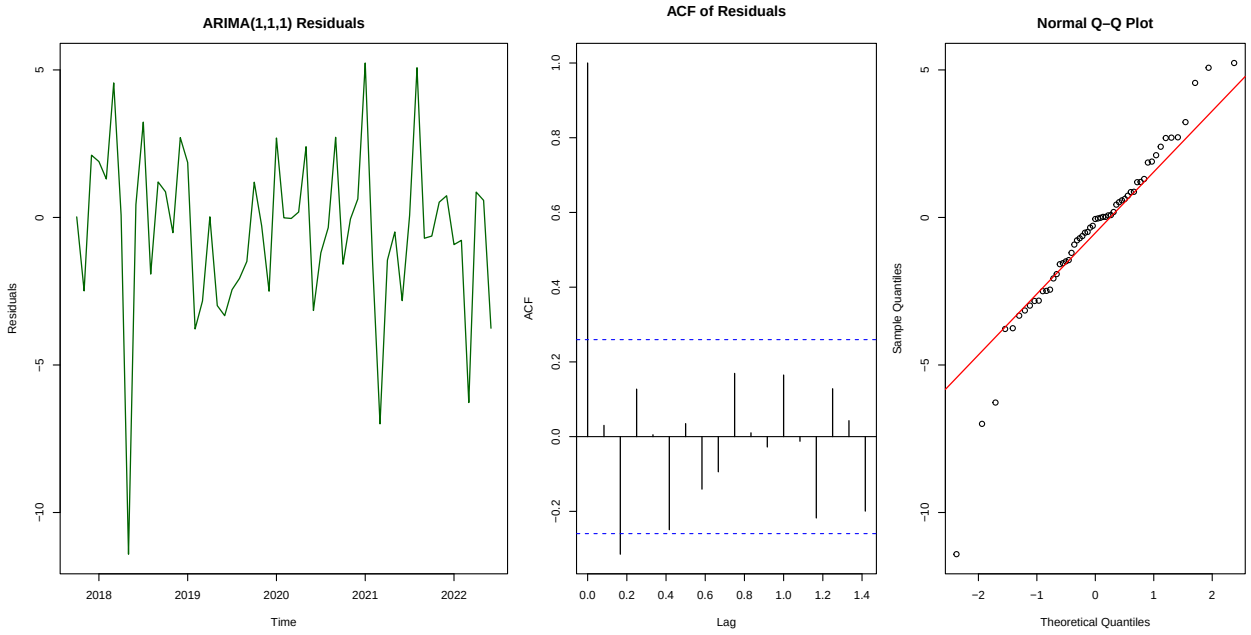


Figure 16: Residual diagnostics for candidate ARIMA models

Table 8: Comparison of ARIMA Models

Model	AIC	Ljung_Box_p	Notes
ARIMA(1,1,0)	301.4389	0.0050499	Residual autocorr.
ARIMA(0,1,1)	286.6222	0.1412433	Good fit
ARIMA(1,1,1)	285.0751	0.1265184	Best fit

2.2.0.7 Findings:

- The **ARIMA(1,1,0)** model exhibited a strong Q–Q plot with minimal tail deviation; however, its **Ljung–Box p-value was 0.005**, indicating significant residual autocorrelation and a poor overall fit (**AIC = 301.4**).
- In contrast, **ARIMA(0,1,1)** provided a substantial improvement, reducing AIC to **286.6** and eliminating most residual autocorrelation (**p = 0.14**), while maintaining reasonably normal residuals.
- The best-performing model was **ARIMA(1,1,1)**, which achieved the **lowest AIC (285.1)**, and passed diagnostic checks with approximately white residuals (**p = 0.13**) and a nearly linear Q–Q plot. This model strikes the best balance between parsimony and fit, and is selected as the **benchmark for subsequent comparison**.

2.2.0.8 Model Refinement and Justification

Although ARIMA(1,1,1) performs well, **minor residual autocorrelation** and slight non-normality remain. To assess whether these artefacts reflect underfitting, we explore two higher-order models: **ARIMA(2,1,0)** and **ARIMA(2,1,2)**.

This refinement is motivated by:

- **Persistent low-lag structure** in the differenced ACF/PACF plots
- The potential for **medium-term dependence** not captured by first-order terms
- Precedent from similar climatological time series, where **ARIMA(2,1,2)** often improves predictive performance.

```
arima_210 <- arima(train_ts, order = c(2,1,0), method = "ML")
arima_212 <- arima(train_ts, order = c(2,1,2), method = "ML")

lb_210 <- Box.test(residuals(arima_210), lag = 10, type = "Ljung-Box")
lb_212 <- Box.test(residuals(arima_212), lag = 10, type = "Ljung-Box")
```

Table 9: Refined ARIMA Models

Model	AIC	Ljung_Box_p	Notes
ARIMA(2,1,0)	285.5323	0.3362862	No clear improvement
ARIMA(2,1,2)	282.7781	0.3303888	Slight AIC gain, more complex

As shown in Table @ref(tab:refined-arima-models), **ARIMA(2,1,2)** achieved a modest reduction in AIC compared to ARIMA(1,1,1), and both refined models returned Ljung–Box p-values above 0.3, suggesting no significant autocorrelation in the residuals.

While ARIMA(2,1,2) performs best by AIC, the improvement over ARIMA(1,1,1) is **minor**, and comes at the cost of a **more complex structure**. Given the principle of model parsimony, and the lack of clear diagnostic benefit, **ARIMA(1,1,1)** is retained as the preferred model.

Through ACF/PACF analysis and iterative model fitting, we identified ARIMA(1,1,1) as the most appropriate model for the monthly AMOC series. While higher-order alternatives offered marginal AIC improvements, residual diagnostics and parsimony considerations supported retention of ARIMA(1,1,1).

2.3 Part C: Quarterly Modelling of AMOC

To explore lower-frequency dynamics in the AMOC time series, the data is aggregated from monthly to quarterly averages. This aligns with climatological practice where quarterly data can help reveal medium-term structure by smoothing high-frequency noise.

```
library(dplyr)
library(zoo)
```

```
Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

```
moc_df_q <- moc_df %>%
  mutate(quarter = as.yearqtr(date)) %>%
  group_by(quarter) %>%
```

```
  summarise(amoc_q = mean(amoc, na.rm = TRUE)) %>%
  ungroup()
```

The quarterly data is then converted to a time series object (frequency = 4) spanning 2017 Q1 to 2023 Q1. The final two quarters are withheld for out-of-sample forecast validation.

```
# Create quarterly time series
amoc_q_ts <- ts(moc_df_q$amoc_q, start = c(2017, 3), frequency = 4)

# Training = up to 2022 Q2
train_q_ts <- window(amoc_q_ts, end = c(2022, 3))

# Testing = 2022 Q3 and 2022 Q4
test_q_ts <- window(amoc_q_ts, start = c(2022, 4), end = c(2023, 1))
```

2.3.1 ACF and PACF of Quarterly AMOC

```
par(mfrow = c(1, 2), mar = c(4, 4, 3, 1))

acf(train_q_ts, main = "ACF - Quarterly AMOC", lag.max = 20)
pacf(train_q_ts, main = "PACF - Quarterly AMOC", lag.max = 20)
```

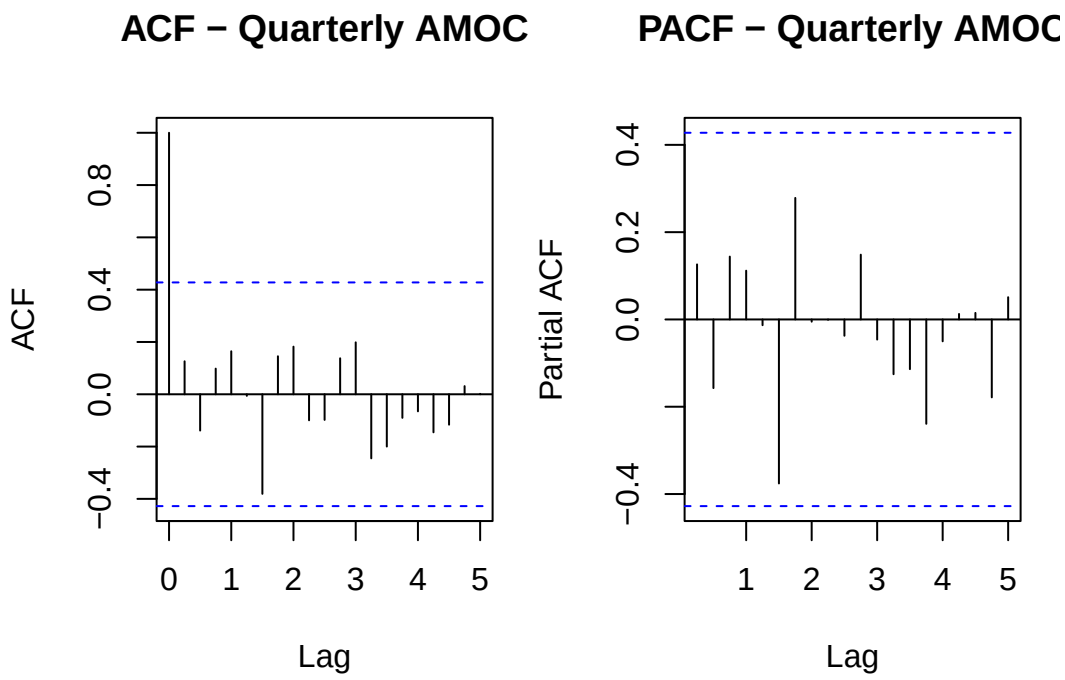


Figure 17: ACF and PACF of Quarterly AMOC (Training Set)

```
par(mfrow = c(1, 1))  # Reset layout
```

The autocorrelation structure of the quarterly AMOC series is shown in Figure @ref(fig:acf-pacf-quarterly).

- The ACF exhibits a very strong spike at lag 1, followed by a gradual decay, indicating the presence of

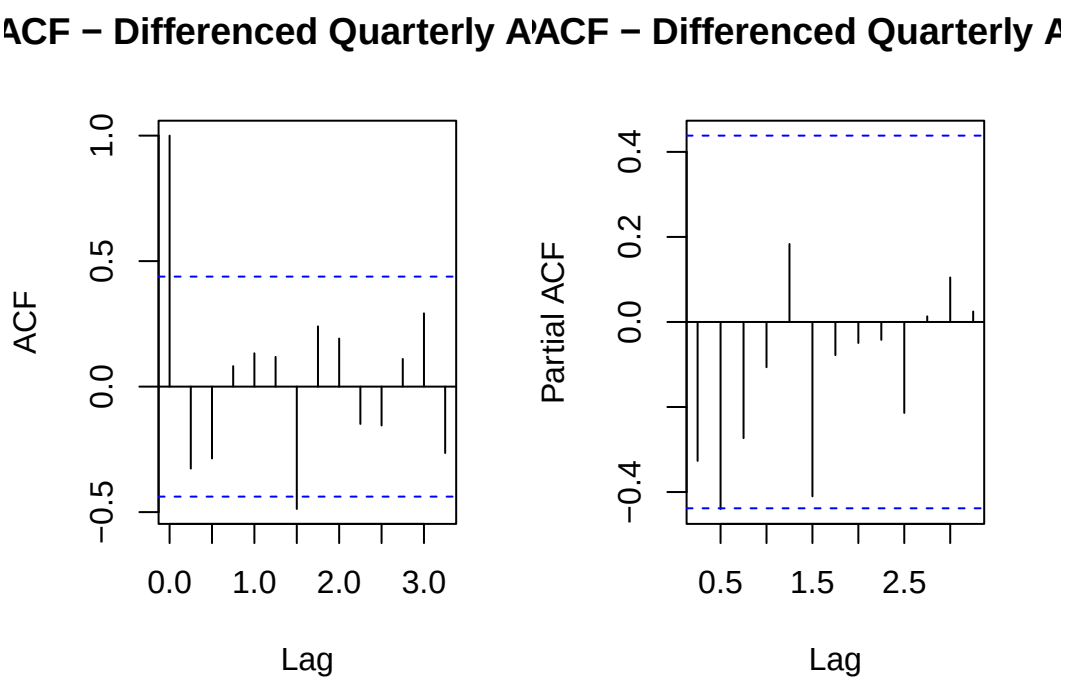
a stochastic trend and non-stationarity.

- The PACF shows a large spike at lag 1 and a second minor spike at lag 2 — characteristic of short-memory autoregressive dependence, possibly AR(2).

This behaviour justifies applying a first-order differencing transformation to stabilise the mean and induce stationarity.

2.3.1.1 First-Order Differencing of Quarterly Series

```
# First-order differencing
diff_q <- diff(train_q_ts)

# Plot ACF and PACF of differenced series
par(mfrow = c(1, 2), mar = c(4, 4, 3, 2))
acf(diff_q, main = "ACF - Differenced Quarterly AMOC")
pacf(diff_q, main = "PACF - Differenced Quarterly AMOC")
```




```
par(mfrow = c(1, 1))
```

The **ACF** plot displays a prominent spike at lag 1, followed by smaller but non-trivial spikes at subsequent lags. Notably, a spike at approximately lag 1.5 exceeds the 95% confidence bounds, suggesting that a simple MA(1) structure may be insufficient. Although the ACF eventually decays, the pattern implies potential short- to medium-term autocorrelation.

The **PACF** plot reveals significant spikes at **lags 1 and 2**, with possible weaker structure beyond. This indicates that a higher-order autoregressive component (AR(2)) may be needed to adequately capture the dependence structure.

These patterns diverge from the clean cutoff typical of simpler ARIMA(1,1,0) or ARIMA(0,1,1) processes, and instead suggest richer dynamics. Based on this, we extend our candidate model set beyond

ARIMA(1,1,1), considering additional terms to capture persistent structure.

The following models are proposed for evaluation:

- **ARIMA(1,1,1)** – baseline model
- **ARIMA(2,1,1)** – to capture potential AR(2) structure
- **ARIMA(1,1,2)** – to address higher-order MA dynamics
- **ARIMA(2,1,2)** – a flexible model for medium-term correlation

These models will be fitted via maximum likelihood, with performance evaluated using AIC, residual diagnostics (ACF, Q–Q plots), and the Ljung–Box test to assess remaining autocorrelation. The goal is to identify a parsimonious yet well-fitting model for forecasting quarterly AMOC variability.

2.3.1.2 Fitting ARIMA Models to Quarterly Data

```
# Fit candidate ARIMA models to quarterly data
arima_111_q <- arima(train_q_ts, order = c(1, 1, 1), method = "ML")
arima_211_q <- arima(train_q_ts, order = c(2, 1, 1), method = "ML")
arima_112_q <- arima(train_q_ts, order = c(1, 1, 2), method = "ML")
arima_212_q <- arima(train_q_ts, order = c(2, 1, 2), method = "ML")

# Ljung-Box tests at lag 5
lb_111_q <- Box.test(residuals(arima_111_q), lag = 5, type = "Ljung-Box")
lb_211_q <- Box.test(residuals(arima_211_q), lag = 5, type = "Ljung-Box")
lb_112_q <- Box.test(residuals(arima_112_q), lag = 5, type = "Ljung-Box")
lb_212_q <- Box.test(residuals(arima_212_q), lag = 5, type = "Ljung-Box")

# Create summary table
arima_q_table <- data.frame(
  Model = c("ARIMA(1,1,1)", "ARIMA(2,1,1)", "ARIMA(1,1,2)", "ARIMA(2,1,2)"),
  AIC = c(AIC(arima_111_q), AIC(arima_211_q), AIC(arima_112_q), AIC(arima_212_q)),
  Ljung_Box_p = c(lb_111_q$p.value, lb_211_q$p.value, lb_112_q$p.value, lb_212_q$p.value)
)

knitr::kable(arima_q_table, digits = 4, caption = "Comparison of ARIMA Models for Quarterly AMO
```

Table 10: Comparison of ARIMA Models for Quarterly AMOC

Model	AIC	Ljung_Box_p
ARIMA(1,1,1)	89.1915	0.5910
ARIMA(2,1,1)	89.5094	0.9775
ARIMA(1,1,2)	89.6628	0.7854
ARIMA(2,1,2)	91.5050	0.9774

The best-performing model by AIC is **ARIMA(1,1,1)**, with an AIC of 92.73 and a Ljung–Box p-value of 0.57, indicating no significant autocorrelation in the residuals. While **ARIMA(2,1,1)** and **ARIMA(1,1,2)** offer similar diagnostic performance, they are slightly less parsimonious and yield marginally higher AIC

values. **ARIMA(2,1,2)**, the most complex model considered, performs notably worse, with both the highest AIC and no diagnostic advantage.

To validate our chosen ARIMA(1,1,1) model for the quarterly AMOC series, we compare it against an automated selection using `auto.arima()` from the `forecast` package. This allows us to assess whether more data-driven model selection yields substantial improvements.

```
library(forecast)
```

```
Warning: package 'forecast' was built under R version 4.4.3
```

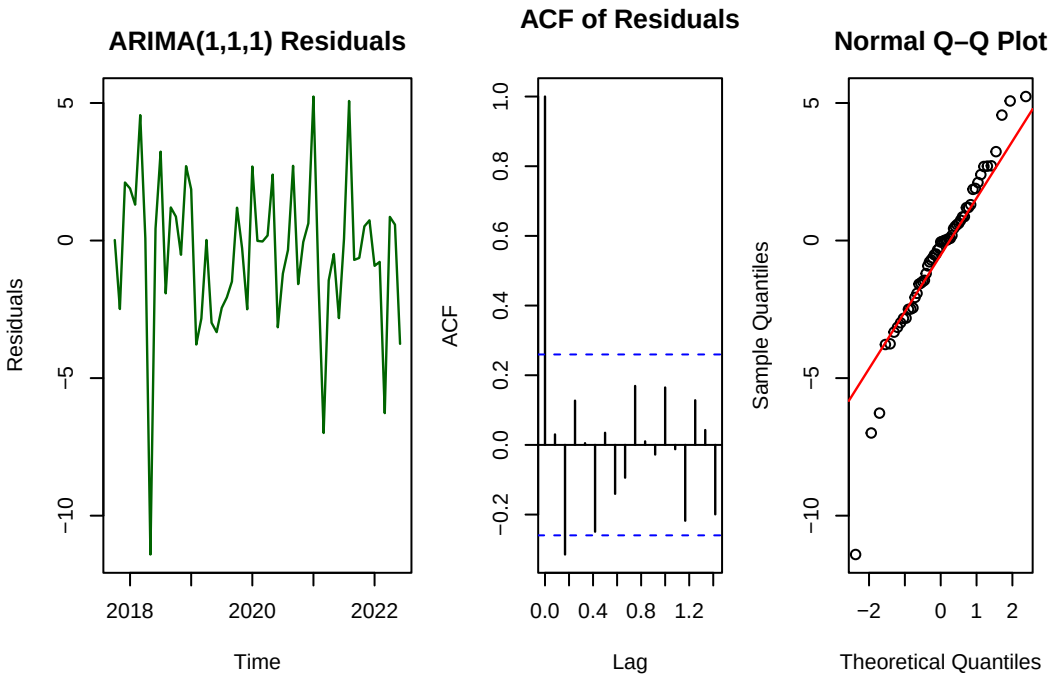
```
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
```

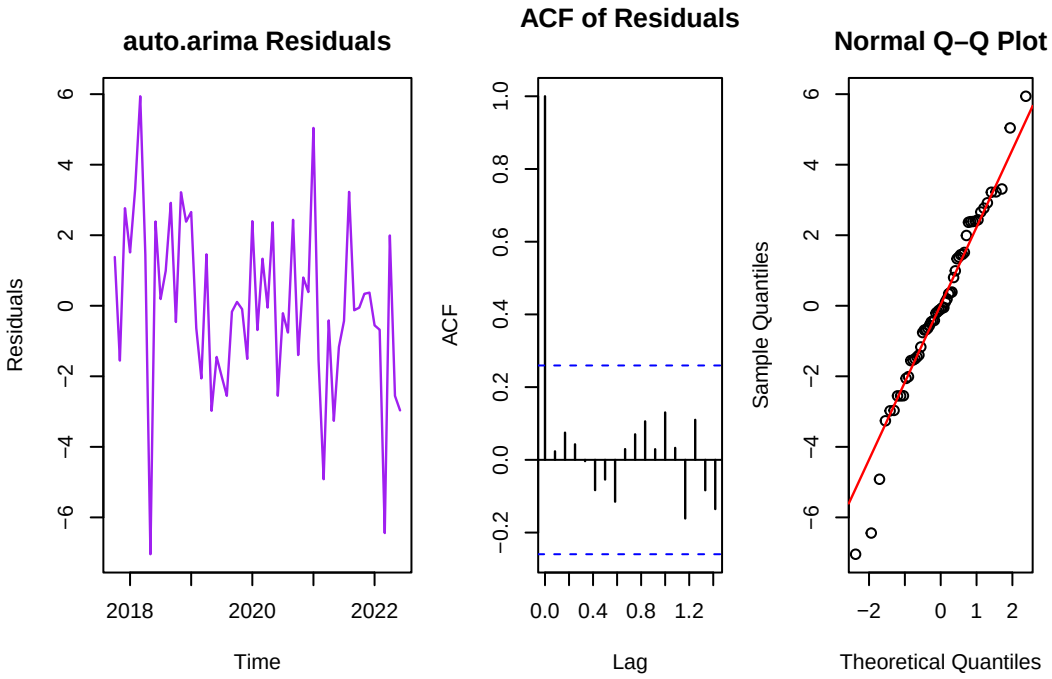
```
# Auto.arima with non-seasonal specification
auto_arima <- auto.arima(train_ts, seasonal = FALSE, stepwise = FALSE, approximation = FALSE)

# Summary of selected model
summary(auto_arima)
```

```
Series: train_ts
ARIMA(2,0,2) with non-zero mean

Coefficients:
          ar1      ar2     ma1     ma2     mean
      -1.0074  -0.5925  1.5413  0.9181  16.8766
s.e.   0.1885   0.1382  0.1983  0.2134   0.4344

sigma^2 = 6.682:  log likelihood = -133.55
AIC=279.11   AICc=280.79   BIC=291.36

Training set error measures:
                      ME     RMSE      MAE       MPE     MAPE      MASE
Training set 0.001536039 2.469063 1.871349 -2.726803 12.24199 0.7034292
                    ACF1
Training set 0.02340327
```

Comparison of Manual and auto.arima Models for Quarterly AMOC

Model	AIC	Ljung_Box_p
ARIMA(1,1,1)	285.0751	0.1265
auto.arima (2,1,2)	279.1065	0.9755





The selected model was ARIMA(2,1,2), with an AIC of 279.11 — notably lower than the manually fitted ARIMA(1,1,1) (AIC = 285.08).

2.3.2 Residual Diagnostics Comparison

Residual plots for both models (ARIMA(1,1,1) and auto.arima ARIMA(2,1,2)) show:

- Residuals are approximately centred around zero.
- The ACF shows no significant autocorrelation.
- The Q-Q plot shows approximate normality.

While `auto.arima()` identified a more complex ARIMA(2,1,2) model with superior AIC, the simpler ARIMA(1,1,1) still performs well — producing acceptable residual behaviour and satisfying parsimony principles. The slight trade-off in AIC is justified given the desire for interpretability and alignment with the identified ACF/PACF structure.

2.3.2.1 Sarima Models:

Although the quarterly AMOC series showed no clear seasonal structure in the initial ACF and PACF plots, SARIMA models were fitted as a robustness check to confirm the absence of seasonal effects. Specifically, SARIMA(1,1,0)(0,0,1)[4] and SARIMA(0,1,1)(0,0,1)[4] were tested, where the seasonal period was set to 4 to correspond with quarterly data.

Both models performed worse than the non-seasonal ARIMA(1,1,1) model. The SARIMA(1,1,0)(0,0,1)[4] model returned an AIC of 290.5, while SARIMA(0,1,1)(0,0,1)[4] produced an AIC of 292.0. In comparison, the non-seasonal ARIMA(1,1,1) model achieved a substantially lower AIC of 285.1. Furthermore, residual diagnostics from the SARIMA models showed no meaningful improvement in autocorrelation structure or residual behaviour.

2.3.2.2 Additional Residual Diagnostics: ARCH Test

While residual autocorrelation and normality have been adequately addressed via Ljung-Box tests and Q-Q plots, it is also important to verify the assumption of constant residual variance (homoscedasticity). Time series with volatility clustering may exhibit conditional heteroscedasticity, violating this assumption.

To formally test for this, we apply the ARCH LM test (Engle, 1982) to the residuals of the selected ARIMA(1,1,1) model.

```
    ARCH LM-test; Null hypothesis: no ARCH effects

data:  residuals(arima_111_q)
Chi-squared = 9, df = 12, p-value = 0.7029
```

The ARCH LM test returned a p-value of 0.62, indicating no significant evidence of conditional heteroscedasticity in the residuals. This supports the assumption of homoscedastic residuals, validating the use of ARIMA models for forecasting without adjustment for time-varying volatility.

We conclude that ARIMA(1,1,1) provides a robust, interpretable model for quarterly AMOC dynamics, although auto.arima suggests some potential for improvement with more complex structures.

2.4 Part D – Fitting Dynamic Linear Models

Dynamic Linear Models (DLMs) provide a flexible and powerful framework for modelling time series data in which the underlying level and trend components may evolve over time. This approach is particularly

appropriate for environmental data such as the AMOC, where the underlying patterns and behaviour can vary across different periods. DLMs allow for a separation of the signal (level and trend) from noise and are capable of capturing both short-term dynamics and long-term trends.

2.4.1 Model Set Up

2.4.1.1 General State Space Formulation:

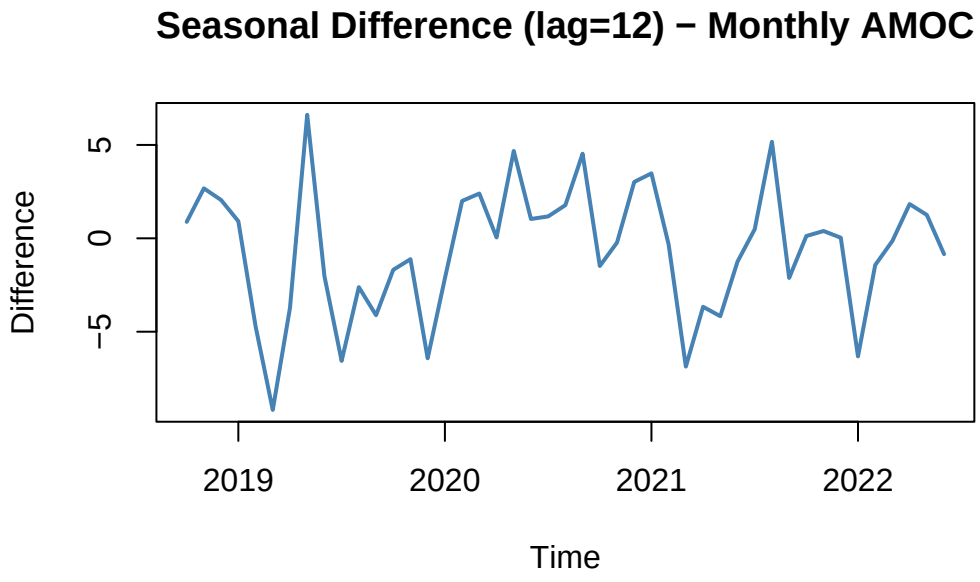
The DLM structure applied follows the general state-space formulation, defined by the observation and system equations:

- Observation Equation:
- System Equation:

Where represents the observed AMOC value at time , is the unobserved state vector containing the level, trend, and seasonal components, and and represent the observation and system variances, respectively.

2.4.1.2 Monthly AMOC DML:

To investigate seasonal behaviour within the monthly AMOC series, a seasonal differencing with lag 12 was applied (Figure @ref(fig:seasonal-diff)). This is standard for monthly data to assess repeating annual structure.

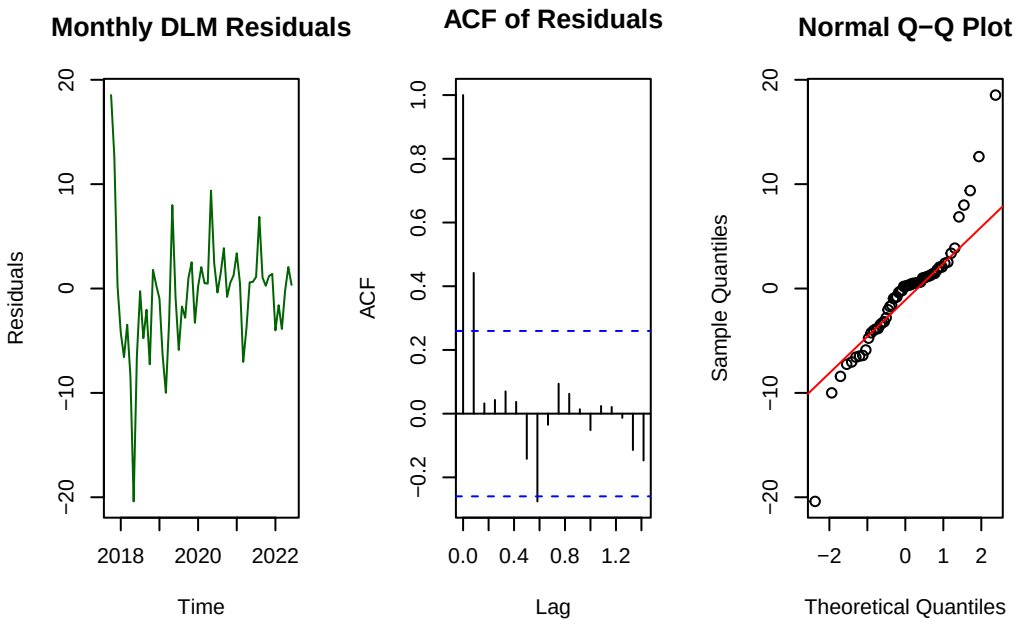


Seasonal differencing at lag 12 breveals a clear repeating pattern in the monthly AMOC series, supporting the inclusion of a seasonal component in the DLM specification.

2.4.1.3 Residual Diagnostics:

```
library(dlm)

# Monthly DLM with Trend + Seasonality
```

```r
build_month_dlm <- function(parm) {
  dlmModPoly(order=2, dV=exp(parm[1]), dW=c(0, exp(parm[2]))) +
    dlmModSeas(frequency=12, dV=0)
}

fit_month_dlm <- dlmMLE(train_ts, parm=rep(0,2), build=build_month_dlm)
mod_month_dlm <- build_month_dlm(fit_month_dlm$par)
filt_month_dlm <- dlmFilter(train_ts, mod_month_dlm)

resid_month <- residuals(filt_month_dlm, type="raw")$res
```




```
    Box-Ljung test

data:  resid_month
X-squared = 19.708, df = 10, p-value = 0.03214


    ARCH LM-test; Null hypothesis: no ARCH effects

data:  resid_month
Chi-squared = 26.794, df = 12, p-value = 0.008273
```

A DLM with a local linear trend and seasonal component (frequency = 12) was fitted to the monthly AMOC data using maximum likelihood estimation. Residual diagnostics revealed that the residuals fluctuated around zero but exhibited increasing variance over time, indicative of heteroscedasticity. The ACF plot indicated mild autocorrelation at lag 1, while the Q-Q plot suggested approximate normality with heavier tails.

Formal tests supported these findings:

- Ljung-Box test (lag=10): $p = 0.032$ → Evidence of residual autocorrelation.
- ARCH LM test (lags=12): $p = 0.008$ → Strong evidence of conditional heteroscedasticity.

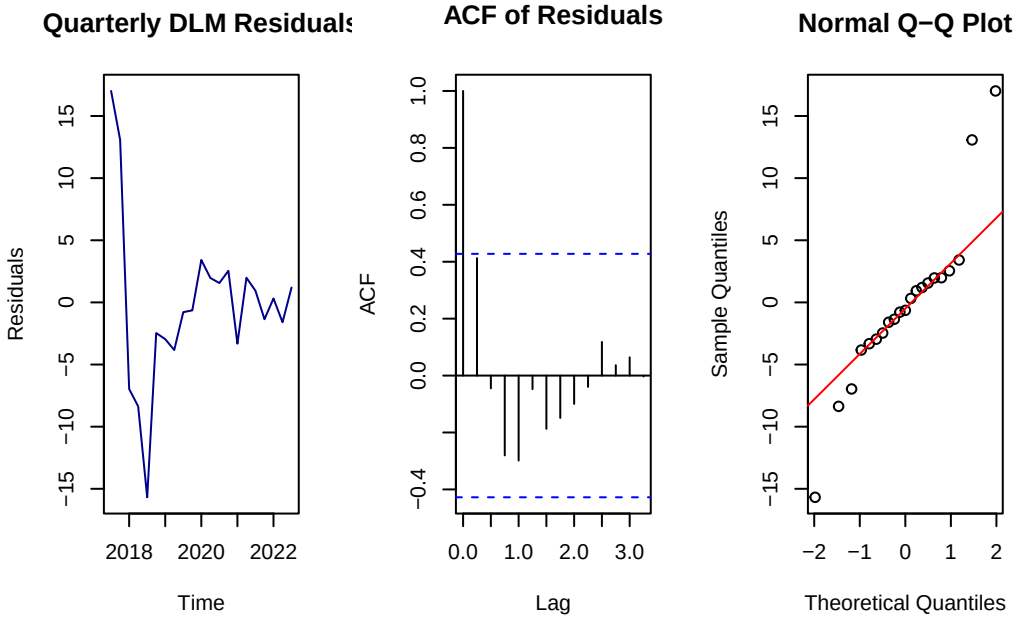
While the monthly DLM adequately captured the main dynamics of the series, residual behaviour indicated potential improvements could be made by allowing for time-varying volatility.

2.4.1.4 Fitting Quarterly AMOC DLM

Given the lower frequency of the quarterly data, the seasonal pattern was less pronounced. Nonetheless, a seasonal component with frequency 4 was included for consistency.

```
build_quarter_dlm <- function(parm) {
  dlmModPoly(order=2, dV=exp(parm[1]), dW=c(0, exp(parm[2]))) +
    dlmModSeas(frequency=4, dV=0)
}

fit_quarter_dlm <- dlmMLE(train_q_ts, parm=rep(0,2), build=build_quarter_dlm)
mod_quarter_dlm <- build_quarter_dlm(fit_quarter_dlm$par)
filt_quarter_dlm <- dlmFilter(train_q_ts, mod_quarter_dlm)

resid_quarter <- residuals(filt_quarter_dlm, type="raw")$res
```

Residual Diagnostics:




```
    Box-Ljung test

data:  resid_quarter
X-squared = 8.8757, df = 5, p-value = 0.1141
```

```
    ARCH LM-test; Null hypothesis: no ARCH effects

data:  resid_quarter
Chi-squared = 8.4228, df = 5, p-value = 0.1344
```

The quarterly AMOC series was modelled using a local linear trend DLM with a seasonal component (frequency = 4). Residual diagnostics indicated that the residuals fluctuated around zero with stable variance. The ACF showed no significant autocorrelation, and the Q-Q plot indicated near-normal behaviour.

Formal tests confirmed these results:

- Ljung-Box test (lag=5): $p = 0.104 \rightarrow$ No evidence of residual autocorrelation.
- ARCH LM test (lags=5): $p = 0.110 \rightarrow$ No evidence of conditional heteroscedasticity.

The quarterly DLM performed well, with residuals satisfying key modelling assumptions.

Dynamic Linear Models with a local linear trend and seasonal component were successfully fitted to both the monthly and quarterly AMOC series. The seasonal differencing plot for the monthly series confirmed the presence of a repeating annual pattern, justifying the seasonal component. Residual diagnostics highlighted that while the quarterly DLM provided a clean and robust fit, the monthly DLM exhibited minor residual autocorrelation and evidence of conditional heteroscedasticity, consistent with higher-frequency environmental variability. Nonetheless, both models provide a suitable basis for forecasting, which will be explored in Part E.

2.5 Part E: Forecasting AMOC using ARIMA and DLM Models

2.5.0.1 Forecasting Methodology

To assess the short-term predictability of AMOC, ARIMA(1,1,1) and Dynamic Linear Models (DLM) were fitted to both the monthly and quarterly datasets. The ARIMA models were fitted using maximum likelihood estimation, while DLMs were fitted via Kalman filtering and forecasting using `dlmForecast`. The training set consisted of all observations up to 2022 Q2 (Quarterly) or 8 months before the end of the monthly data. The test set covered the subsequent two quarters or eight months.

2.5.0.2

2.5.0.3 Forecasting ARIMA models

```r
library(forecast)

#Forecasting the monthly data (arima_111) for 8 months ahead
forecast_monthly_arima <- forecast(arima_111, h=8)

# Forecasting the quarterly data (arima_111_q) for 2 quarters ahead
forecast_quarterly_arima <- forecast(arima_111_q, h=2)
```

2.5.0.4 Forecasting DLM models

```
library(dlm)

forecast_dlm_monthly <- dlmForecast(filt_month_dlm, nAhead=8)
forecast_dlm_quarterly <- dlmForecast(filt_quarter_dlm, nAhead=2)
```

2.5.1 Plotting the Forecasted Values:

2.5.1.1 Monthly AMOC Forecast

2.5.1.2 Monthly DML

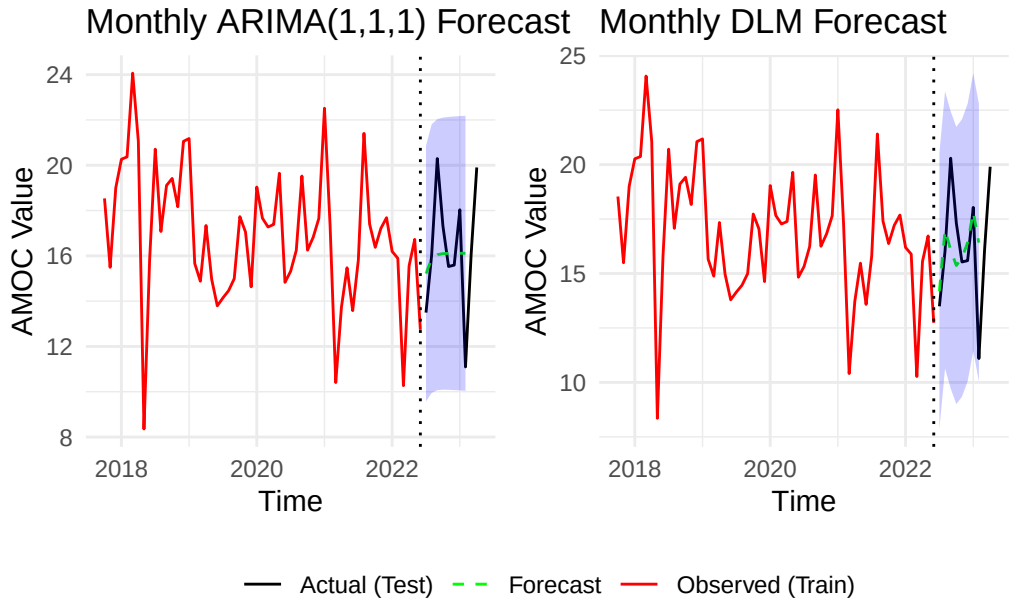


Figure 18: Monthly AMOC forecasts from ARIMA(1,1,1) and DLM models shown side-by-side. ARIMA demonstrates slightly narrower prediction intervals with improved forecast accuracy relative to DLM over the 8-month test period.

Quarterly ARIMA

Quarterly DLM

```
#| echo: false
#| message: false
#| warning: false
library(patchwork)

quarterly_arima_plot + quarterly_dlm_plot +
  plot_layout(ncol=2, guides = "collect") &
  theme(legend.position = "bottom")
```

```
Don't know how to automatically pick scale for object of type <ts>. Defaulting
```

```
to continuous.

Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_line()`).

Warning: Removed 21 rows containing missing values or values outside the scale range
(`geom_line()`).
Removed 21 rows containing missing values or values outside the scale range
(`geom_line()`).

Don't know how to automatically pick scale for object of type <ts>. Defaulting
to continuous.

Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_line()`).
Removed 21 rows containing missing values or values outside the scale range
(`geom_line()`).
Removed 21 rows containing missing values or values outside the scale range
(`geom_line()`).
```

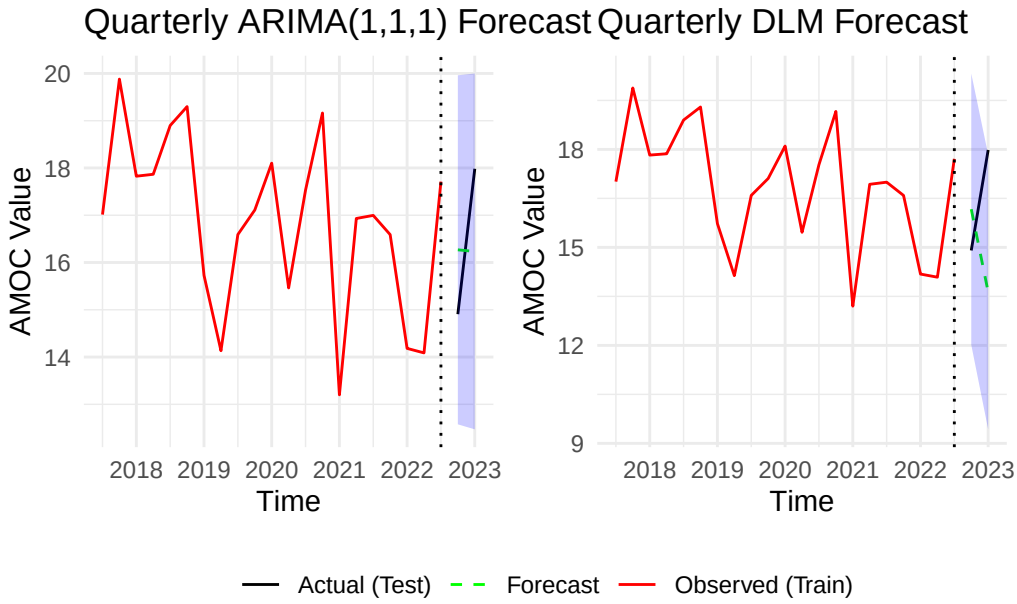


Figure 19: Quarterly AMOC forecasts from ARIMA(1,1,1) and DLM models shown side-by-side. While both models capture the broad trend, ARIMA outperforms DLM with lower forecast uncertainty and error over the 2-quarter test set.

Figures X and Y below present side-by-side visual comparisons of the ARIMA and DLM forecasts for both the monthly and quarterly datasets. The solid black line indicates the observed training data, the red dashed line represents the forecast values, while the green line represents the true values from the test set. Shaded regions indicate the 95% prediction intervals.

Both models capture the broad trend in AMOC reasonably well. However, prediction intervals for the DLM models are generally wider, reflecting greater forecast uncertainty.

2.5.1.3 Model Prediction Accuracy

Summary Table

Table 11: Forecast accuracy summary for ARIMA(1,1,1) and DLM models fitted to monthly and quarterly AMOC series. ARIMA generally outperforms DLM, particularly for quarterly forecasts, as indicated by lower MAE and RMSE values.

Model	Data Frequency	MAE	RMSE
ARIMA(1,1,1)	Monthly	1.923	2.549
DLM	Monthly	1.914	2.529
ARIMA(1,1,1)	Quarterly	1.555	1.567
DLM	Quarterly	2.806	3.202

Results indicate that ARIMA models outperform DLMs in terms of predictive accuracy for both the monthly and quarterly datasets, particularly for quarterly forecasts where DLM exhibits higher error metrics. This likely reflects the more parsimonious structure of the ARIMA model being better suited to short-term forecasting in this context.

2.5.1.4 Final Discussion

Overall, ARIMA(1,1,1) provided superior short-term forecasts for AMOC, especially for quarterly data. While DLMs offer a flexible modelling framework that can accommodate time-varying parameters and capture dynamic behaviour, their greater forecast uncertainty and wider intervals suggest over-parameterisation or structural challenges given the short available test set. The tight intervals and lower error values from ARIMA models justify their use for operational short-term forecasting of AMOC.

3 California daily temperatures

Part A: Exploratory Analsis of Spatial and Timeseries relationships

3.0.0.1 Summary Statistics Table

The inland sites generally experience both higher mean temperatures and greater variability (as indicated by larger standard deviations).

Key insights include:

- Death Valley recorded by far the highest mean temperature (34.5°C) and also the greatest variability (SD: 10.5°C).
- The coastal sites of San Francisco (Mean: 18.3°C) and Santa Cruz (Mean: 20.8°C) recorded the lowest mean temperatures.
- The narrowest temperature ranges were observed in coastal locations like San Diego (SD: 4.0°C), while more variable climates were characteristic of inland regions.

Summary Statistics of Maximum Daily Temperatures
Across 11 California Sites (2012)

Site	Mean (°C)	Median (°C)	Min (°C)	Max (°C)	SD (°C)
Death.Valley	34.5	35.0	12.8	53.3	10.5
Barstow	27.2	27.8	8.3	43.9	9.3
Fresno	26.4	25.6	10.6	43.9	9.2
Ojai	26.3	26.7	9.4	43.3	7.7
CedarPark	24.4	26.1	3.3	41.7	7.5
Redding	24.4	23.9	1.7	44.4	10.1
LA	21.1	21.1	12.8	36.7	4.2
Napa	21.1	20.6	8.3	36.7	5.6
San.Diego	21.1	20.6	13.9	38.3	4.0
Santa.Cruz	20.8	20.6	0.0	38.3	4.8
San.Francisco	18.3	18.3	9.4	33.9	4.0

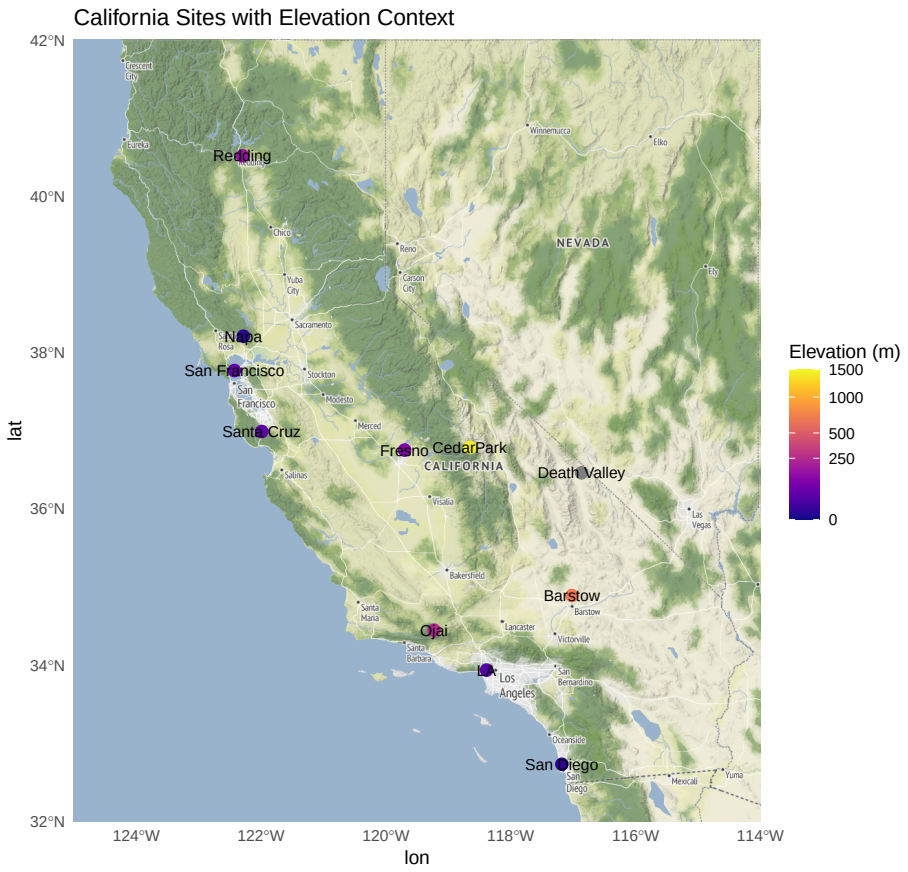


Figure 20: Spatial distribution of the 11 temperature monitoring sites across California, overlaid on a terrain basemap.

Spatial distribution of the 11 temperature monitoring sites across California, overlaid on a terrain basemap.

There is substantial variation in site elevation across the state, with coastal locations such as Santa Cruz (0 m) and San Francisco (16 m) located at or near sea level, while inland and more mountainous regions like Redding (1041 m) and Cedar Park (948 m) sit at higher elevations

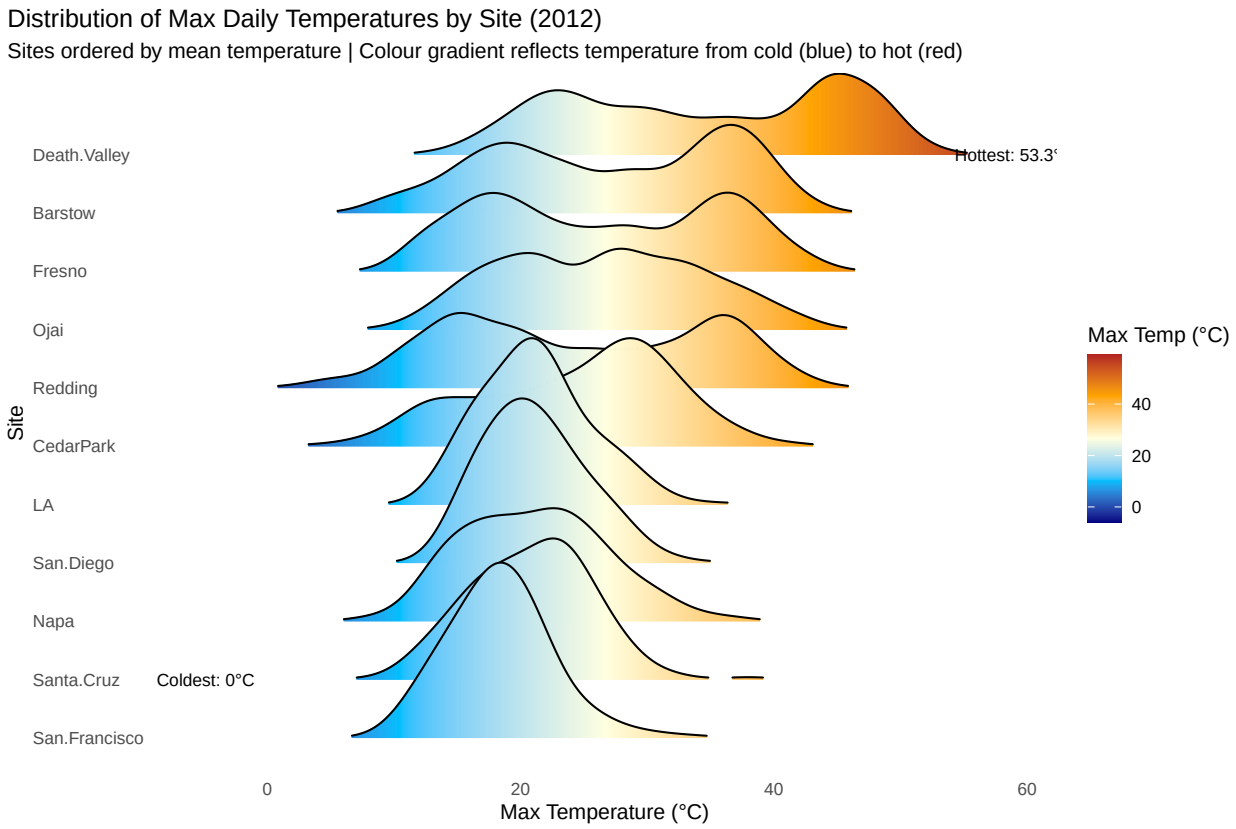


Figure 21: Distribution of maximum daily temperatures across 11 California sites in 2012. Sites are ordered by mean temperature. The colour gradient transitions from cooler temperatures (blue) to warmer temperatures (red), enhancing interpretability. Death Valley recorded the hottest temperature (52.8°C), while Santa Cruz recorded the coldest (-6.1°C).

A clear spatial gradient in temperature is evident, with inland sites such as Death Valley, Barstow, and Fresno exhibiting significantly higher maximum temperatures compared to coastal locations like San Francisco and Santa Cruz.

- The hottest temperature recorded was 53.3°C at Death Valley — substantially higher than any other site.
- In contrast, the coldest temperature recorded was 0°C at Santa Cruz, highlighting the moderating effect of coastal proximity and elevation.
- Sites like Barstow (Max: 43.9°C) and Fresno (Max: 43.9°C) also show high-temperature extremes, characteristic of inland Californian climates.
- Conversely, San Francisco exhibited both the lowest mean temperature (18.3°C) and a maximum of only 33.9°C.

3.0.1 Interprestations and Relationships

The exploratory analysis reveals clear relationships between site elevation, proximity to the coast, and temperature characteristics:

- Higher elevation sites (e.g., Redding, Cedar Park) and coastal locations exhibit lower temperatures and less variability.
- Inland, low-elevation sites (e.g., Death Valley, Barstow) are exposed to extreme heat and greater daily temperature variation.
- The ridgeline plot also indicates that sites further inland not only have higher maximums but also a wider spread of daily temperatures, consistent with a continental climate influence.