A STATISTICAL ANALYSIS OF DAILY EVAPORATION AT CARDINIA RESERVOIR

EXECUTIVE SUMMARY

Melbourne Water Corporation requested an updated statistical model to understand how daily evaporation at Cardinia Reservoir depends on local weather and time-of-year, in order to support operational planning under changing climatic conditions.

Using daily Melbourne weather data over a financial year, I examined how evaporation relates to month, day of week, minimum temperature, maximum temperature and 9 am relative humidity. Bivariate plots showed that evaporation is highest in summer (particularly January–March), lowest in winter (June–August), increases with higher minimum and maximum temperatures, and decreases with higher 9 am humidity. Day of week showed no meaningful pattern.

I then built a multiple linear regression model. Starting from a full model including all candidate predictors and an interaction between month and 9 am humidity, I applied a stepwise backward-selection procedure based on F-tests. Day of week and maximum temperature were not statistically significant once other variables were included and were removed. The final model retained month, minimum temperature, 9 am humidity, and a month × humidity interaction. The model explains around 64% of the variation in daily evaporation.

In this model, higher overnight (minimum) temperatures are associated with higher evaporation, while higher 9 am humidity is associated with lower evaporation. The strength of the humidity effect varies by month, with humidity having a stronger (more negative) impact in some months than others. Summer months (January–March) have higher underlying evaporation than winter months, even after controlling for minimum temperature and humidity.

Model diagnostics showed no strong violations of linear regression assumptions. There is some skewness and heavier-than-normal tails in the residuals, and mild evidence of non-constant variance, but overall the model appears adequate for prediction at daily resolution.

Using the final model, I produced 95% prediction intervals for four specific days in 2020 with specified weather conditions. Under those scenarios, the highest evaporation is expected on 13 January 2020 (around 14.9 mm, 95% prediction interval approximately 10.1–19.6 mm). On that day, we can say with 95% confidence that evaporation will exceed 10 mm and therefore temporary transfer measures would be required. For 29 February and 6 July, the model indicates evaporation will almost certainly be below 10 mm, so temporary measures are not required. For 25 December, the model is inconclusive: the interval spans both sides of 10 mm.

Overall, the analysis confirms that evaporation at Cardinia is primarily driven by season, overnight temperature, and morning humidity, with humidity having a month-dependent effect. The final model can be used to forecast daily evaporation with quantified uncertainty and to identify days on which proactive transfer from Silvan Reservoir is very likely or very unlikely to be needed.

INTRODUCTION

Cardinia Reservoir is a key component of Melbourne's water supply system. Efficient management requires reliable estimates of daily evaporation so that storage volumes and transfers between reservoirs can be planned in advance. Previous evaporation estimates have become less reliable under recent climatic variability, and Melbourne Water Corporation has requested an updated, data-driven model for evaporation based on daily weather observations.

This report uses a full financial year of Melbourne weather data to:

- Identify how daily evaporation depends on temporal and meteorological variables,
- Build and justify a multiple linear regression model predicting evaporation,
- Assess whether the model satisfies standard regression assumptions, and
- Use the model to forecast evaporation under several specific weather scenarios relevant to Cardinia Reservoir operations.

The variables considered as potential predictors are:

- Month of the year,
- Day of the week,
- Minimum temperature (°C),
- Maximum temperature (°C), and
- 9 am relative humidity (%).

The report is structured as follows. The Methods section describes the data, bivariate summaries, and the model-building and diagnostic procedures. The Results section presents the final model and interprets its coefficients. The Discussion interprets prediction intervals for specific days and their operational implications. The report concludes with key recommendations for Melbourne Water Corporation, and the Appendix contains all R code used in the analysis.

METHODS

Data and Variables:

Daily meteorological data for Melbourne over a single financial year were supplied in melbourne.csv. The following variables were used:

- **Response**:
    - evap: daily evaporation in millimetres (Evaporation (mm)).
- **Temporal predictors**:
    - Month: month of the observation (January–December), derived from the Date variable and treated as a categorical factor.
    - Day: day of the week (Monday–Sunday), derived from Date and treated as a categorical factor.
- **Meteorological predictors**:
    - mintemp: minimum daily temperature in °C (Minimum temperature (Deg C)).
    - maxtemp: maximum daily temperature in °C (Maximum Temperature (Deg C)).
    - rh9: 9 am relative humidity in percent (9am relative humidity (%)).

Records with missing evaporation values (8 days) were excluded so that the response was fully observed for modelling. All other variables used in the analysis were complete.

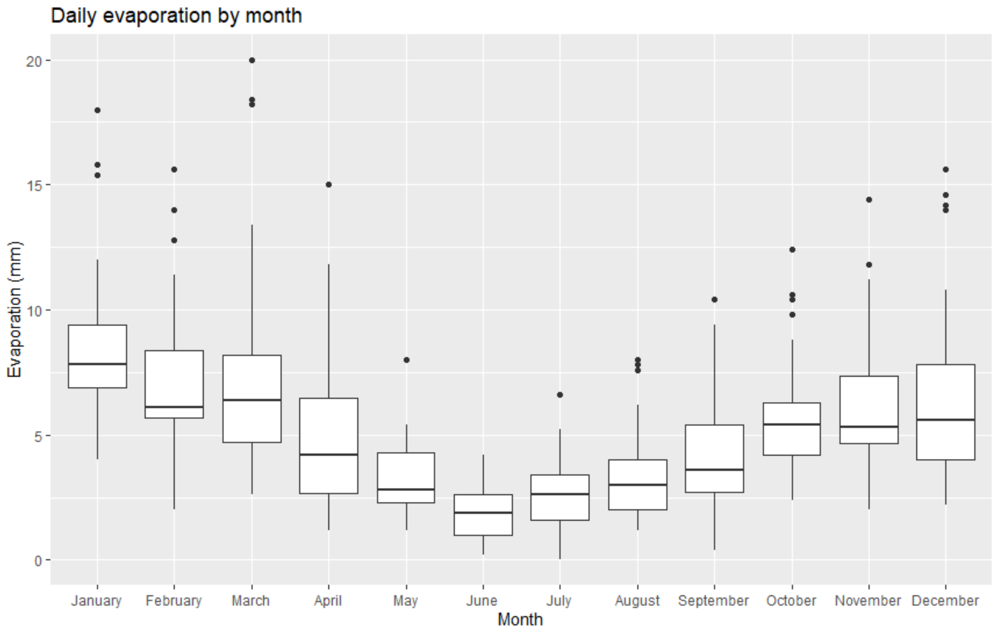
the tidyverse package was used for data wrangling, visualisation and modelling. All R codes are provided in the Appendix.

Bivariate Summaries

To understand the relationship between evaporation and each potential predictor, I produced the following plots:

- **Evaporation vs Month**

A boxplot of evap by month

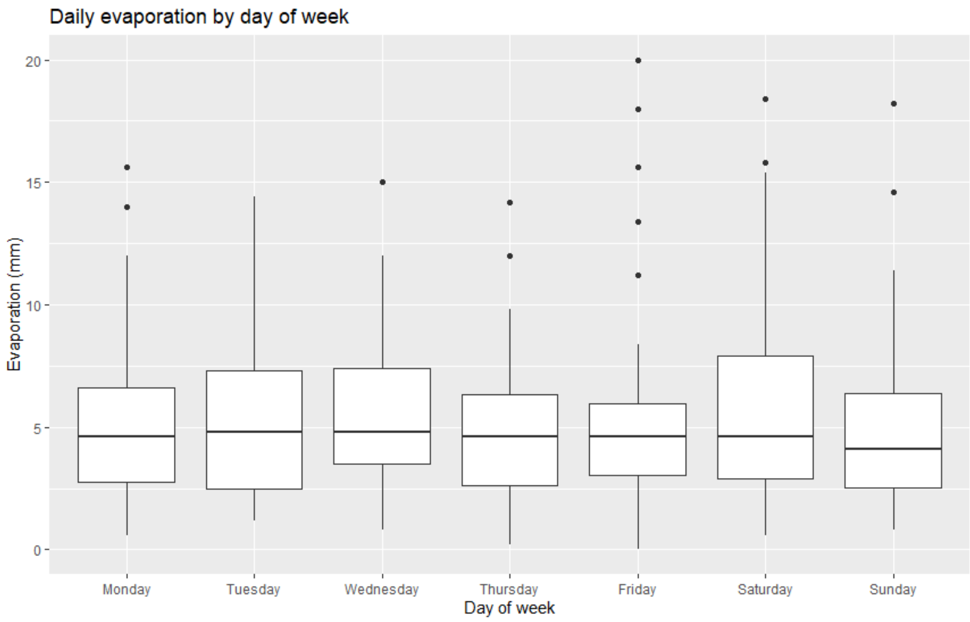


(Figure 1)

This plot shows a clear seasonal pattern: median evaporation is highest in late summer (January–March) and lowest in winter (June–August). Variability is also larger in hotter months.

- **Evaporation vs Day of week**

A boxplot of evap by Day

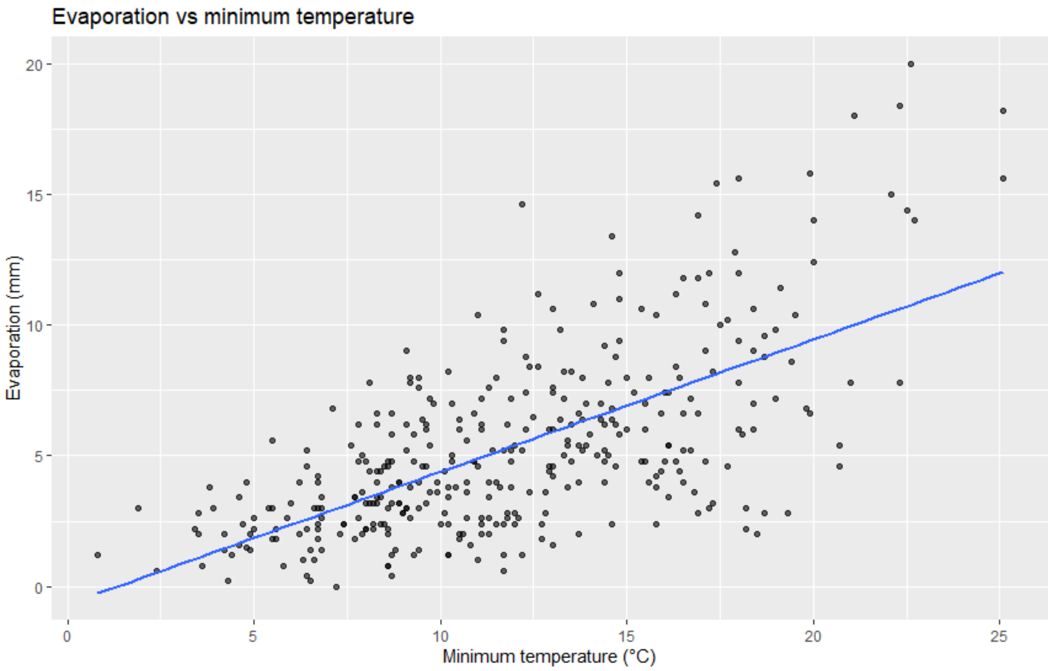


(Figure 2).

Unlike month, evaporation does not show a systematic pattern by day of week; median and spread are broadly similar across Monday–Sunday, as expected for a natural process unrelated to weekdays.

- **Evaporation vs Minimum temperature**

A scatterplot of evap against mintemp with a fitted smooth line.

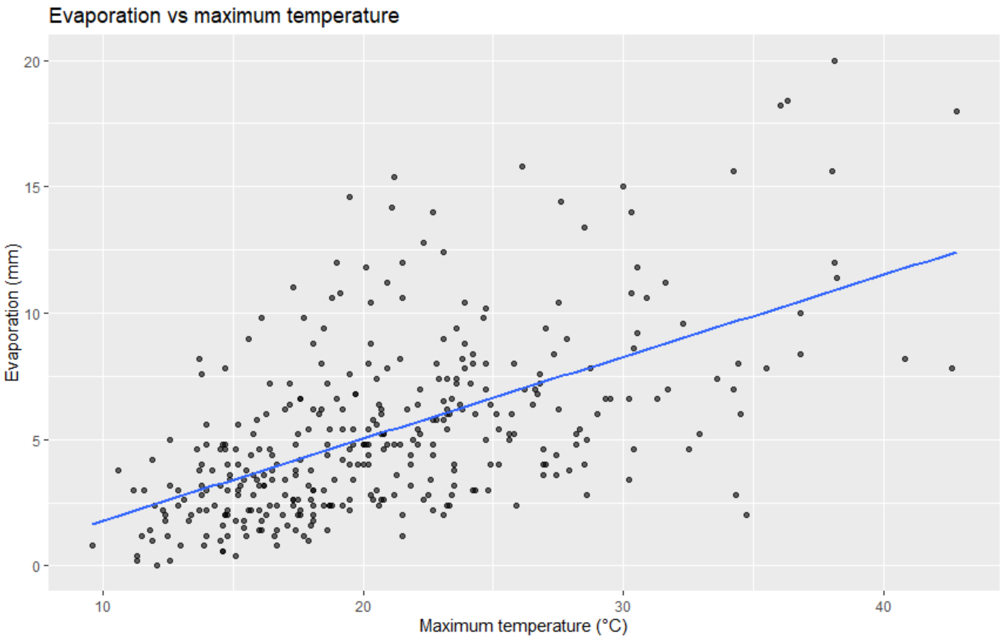


(Figure 3)

There is a clear positive relationship: days with higher overnight minimum temperatures tend to have higher evaporation. The relationship appears roughly linear over the observed range.

- **Evaporation vs Maximum temperature**

A scatterplot of evap against maxtemp with a smooth line.

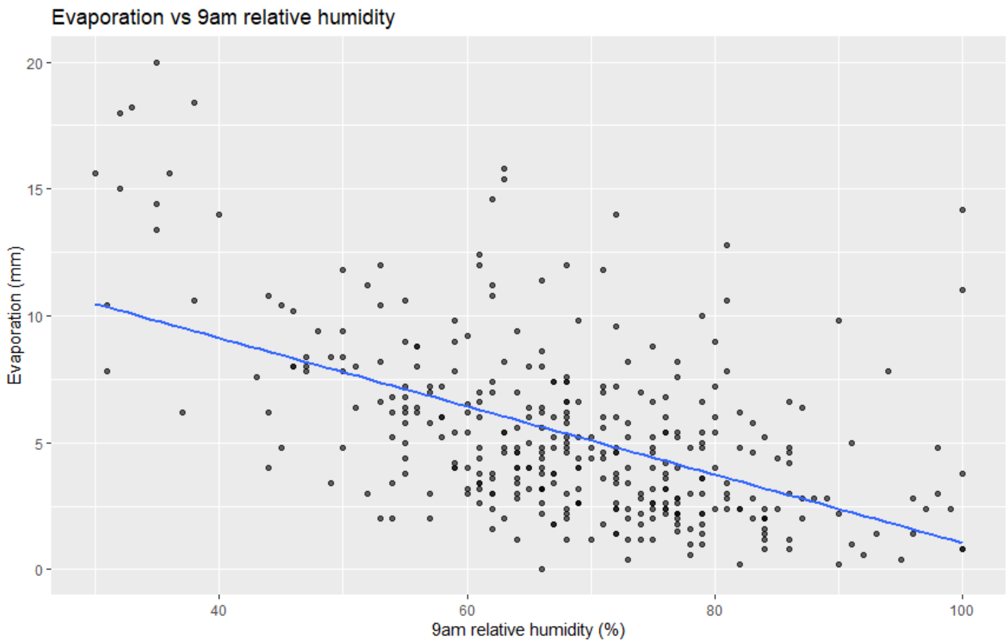


(Figure 4)

Evaporation increases as maximum temperature increases, especially from cool to very hot days. The pattern is generally linear with some flattening at extreme values.

- **Evaporation vs 9 am humidity**

A scatterplot of evap against rh9 with a smooth line.



(Figure 5)

There is a clear negative relationship: higher 9 am humidity is associated with lower evaporation, which is physically plausible because humid air slows evaporation.

Summary

These bivariate plots suggest that month (seasonality), minimum temperature, maximum temperature, and 9 am humidity are all individually associated with evaporation, whereas day of week is not.

Model Building and Selection:

The goal was to build a linear model predicting daily evaporation in millimetres using all five predictors and an interaction between month and 9 am humidity to allow the effect of humidity to vary by month.

The modelling procedure followed the assignment instructions:

1. **Initial Full Model**

I fitted the full model:

$$\text{evap} = \beta_0 + \text{Month} + \text{Day} + \beta_1\,\text{mintemp} + \beta_2\,\text{maxtemp} + \beta_3\,\text{rh9} + (\text{Month} \times \text{rh9}) + \varepsilon.$$

In R:

```
mod0 <- lm(evap ~ Month + Day + mintemp + maxtemp + rh9 + Month:rh9, data =
melb)
```

2. **Assessing Predictor Significance**

- For quantitative predictors (mintemp, maxtemp, rh9), I used the standard regression summary to obtain t-tests and p-values.
- For categorical predictors (Month, Day) and the interaction Month:rh9, I used F-tests from anova/drop1 to test each whole term.

3. **Backward Elimination**

I iteratively removed the least significant term (highest p-value), refitted the model, and repeated the testing until all remaining terms were significant at the 5% level:

- **Day of week**: Testing a model with and without Day showed that the Day of week factor was not significant ($F \approx 1.17$, $p \approx 0.32$). This agrees with the bivariate boxplot and the lack of physical reason for a weekday effect. Day was dropped.
  New model: mod1 <- lm(evap ~ Month + mintemp + maxtemp + rh9 + Month:rh9, data = melb)

- **Interaction term Month:rh9**: Comparing models with and without Month:rh9 showed the interaction was significant ($F \approx 3.23$, $p \approx 0.0003$). This means the effect of humidity on evaporation differs by month, so the interaction was retained.

- **Maximum temperature**: Comparing models with and without maxtemp within the presence of Month, mintemp, rh9 and Month:rh9 showed that maxtemp did not significantly improve model fit ($F \approx 0.14$, $p \approx 0.71$). Once minimum temperature and humidity are included, maximum temperature does not explain additional variation in evaporation. maxtemp was removed.

- **Minimum temperature and humidity**: Both mintemp and rh9 were highly significant ($p \ll 0.001$) and retained. Because of the interaction term, rh9 and Month must remain in the model to preserve hierarchy.

4. **Final Model**

   The final model is:

$$\text{evap} = \beta_0 + \text{Month} + \beta_1 \text{ mintemp} + \beta_2 \text{ rh9} + (\text{Month} \times \text{rh9}) + \varepsilon,$$

Month and Month:rh9 are converted into dummy variables so the model can estimate differences between months and how humidity behaves differently in each month.

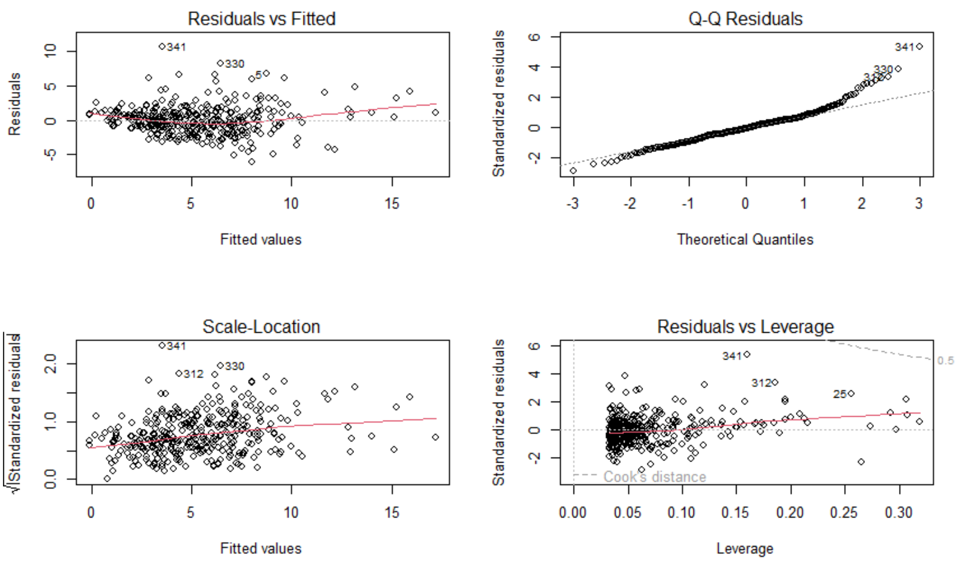
In R:

mod_final <- lm(evap ~ Month + mintemp + rh9 + Month:rh9, data = melb)

This final model has an R-squared of about 0.64, meaning it explains approximately 64% of the variation in daily evaporation in the sample.

Model Diagnostics

To evaluate whether the final model meets the assumptions of linear regression, I produced the standard diagnostic plots for mod_final using plot(mod_final). The plots below were used to assess linearity, variance patterns, normality, independence, and influential points.



(Figure 6)

1. **Linearity and homoscedasticity**
   - The **Residuals vs Fitted** plot shows residuals scattered around zero with no strong curved pattern, indicating that the linear form in mintemp and rh9 is reasonable.
   - There is some increase in spread of residuals at higher fitted values, suggesting mild heteroscedasticity (slightly larger variance on high-evaporation days), but not severe enough to invalidate the model for practical prediction.
2. **Normality of residuals**
   - The **Normal Q–Q plot** shows residuals mostly following the straight line, with some deviations in the upper tail, indicating a slightly right-skewed distribution with heavier tails than a perfect normal distribution. Given the sample size (over 350 days), inference from the model should still be reasonably robust.
3. **Independence**
   - The **Residuals vs Order (or Date)** plot does not show a strong systematic trend over time.
   - The Durbin–Watson statistic is close to 2 (≈ 1.8), indicating only mild positive autocorrelation. For daily data of this type, independence appears acceptable for the purposes of modelling.
4. **Influential points**
   - The **Scale–Location** and **Cook's distance** plots identify a few high-evaporation days as mildly influential, but no single day dominates the model.

Overall, the assumptions of linearity, approximate normality, and constant variance are satisfactorily met for this application, with only mild deviations at extreme evaporation values. The model is therefore suitable for making daily evaporation predictions and associated intervals.

RESULTS

Final Model Summary

In the final model, April serves as the reference month, meaning all month effects are interpreted relative to April. The model includes minimum temperature (mintemp), 9 am relative humidity (rh9), the Month factor, and the Month×rh9 interaction. Month and the interaction are represented in the model using dummy (indicator) variables, allowing evaporation levels and the effect of humidity to vary across months.

Key estimated coefficients (rounded):

- Intercept (April baseline): ≈ 10.56 mm
- Minimum temperature (mintemp): evaporation increases by ≈ 0.37 mm for each 1 °C increase
- 9 am humidity (rh9): evaporation decreases by ≈ 0.15 mm for each 1% increase in humidity (in April)
- Month and Month×rh9 terms: adjust these effects for each month of the year

The full set of estimates is given in the R output (Appendix).

Model Numerical Summary

The final model explains a substantial proportion of the variation in daily evaporation:

- **R^2:** ≈ 0.65
- **Adjusted R^2:** ≈ 0.61
- **Residual standard error:** ≈ 2.18 mm

- **Overall F-statistic:** significant at $p < 2.2 \times 10^{-16}$

These values indicate that the model provides a good fit for the data while avoiding unnecessary predictors.

Interpretation of Coefficients

Because the model includes interactions, the effect of humidity depends on month. Below, I interpret the main effects and provide an example for categorical terms.

Intercept

- The intercept (≈ 10.6 mm) is the predicted evaporation on an April day with minimum temperature and 9 am humidity both equal to zero.
- This combination does not occur in practice but serves as the baseline from which other effects are added or subtracted. The intercept itself is not directly meaningful, but it is necessary to correctly interpret the other coefficients.

Minimum temperature (mintemp)

- For April (the reference month), holding 9 am humidity constant, each increase of 1 °C in minimum temperature is associated with an increase of about 0.37 mm in daily evaporation.
- For example, in April, a day with minimum temperature 15 °C would be expected to evaporate roughly 3.7 mm more water than a day with minimum temperature 5 °C, all else equal.

This reflects the fact that warmer nights (higher minimum temperatures) generally indicate warmer conditions overall, supporting more evaporation.

9 am relative humidity (rh9)

- In April, the coefficient on rh9 is about −0.15 mm per 1% increase in humidity. This means that, in April, if 9 am humidity increases from 50% to 60%, evaporation is predicted to decrease by about 1.5 mm, holding minimum temperature constant.
- However, this “base” humidity effect is modified by the Month × humidity interaction terms.

Month and Month × humidity (example)

- Month shifts the baseline evaporation level relative to April, even at the same minimum temperature and humidity.
- For example, July has a negative main effect relative to April, reflecting consistently lower evaporation in mid-winter.

The interaction allows the effect of humidity to differ by month:

- In January, the humidity effect is:
  base humidity effect + January interaction ≈ −0.15 + 0.05 ≈ −0.10 mm per 1%.
  This means humidity still reduces evaporation in January, but slightly less strongly than in April.
- In June and August, the interaction coefficients are positive and relatively large (around +0.13 and +0.14), partially offsetting the base humidity effect. Humidity still tends to reduce evaporation but with different strength across months.

Overall, the model suggests:

- Evaporation is highest in late summer months (January–March),
- Lowest in winter months (June–August),
- Increases with higher minimum temperatures, and
- Decreases with higher 9 am humidity, with the strength of the humidity effect varying by month.

This final model provides a clear, interpretable structure for predicting daily evaporation at Cardinia Reservoir.

DISCUSSION

Predictions for Specified Days

Melbourne Water Corporation requested predictions for four specific days with given weather conditions. Using the final model, I constructed a prediction dataset with the specified minimum temperature, maximum temperature, and 9 am humidity, and derived month from the given date. Predictions were obtained using 95% prediction intervals, which quantify the range in which the evaporation for an individual day is likely to fall.

Table 1 summarises the results.

Predicted Evaporation and 95% Prediction Intervals for Specified Days

Date	Month	Min temp (°C)	Max temp (°C)	9 am RH (%)	Predicted evap (mm)	95% prediction interval (mm)
29 Feb 2020	February	13.8	23.2	74	5.5	1.1 to 9.9
25 Dec 2020	December	16.4	31.9	57	8.6	4.2 to 13.0
13 Jan 2020	January	26.5	44.3	35	14.9	10.1 to 19.6
6 Jul 2020	July	6.8	10.6	76	2.3	−2.1 to 6.6

(Table 1)

(Intervals are rounded to two decimal places; negative lower bounds in winter should be interpreted practically as zero evaporation.)

Interpretation in context

- **29 February 2020 (late summer, mild day, high humidity)**

  On a relatively mild and humid summer day, the model predicts moderate evaporation of around 5–6 mm, with a 95% prediction interval from approximately 1–10 mm. This reflects the dampening effect of high humidity on evaporation even in summer.

- **25 December 2020 (hot summer day, moderate humidity)**
  On a warmer day in December with moderate humidity, predicted evaporation increases to around 8–9 mm, with a wider 95% prediction interval (about 4–13 mm). Evaporation is more variable under these conditions.

- **13 January 2020 (very hot day, low humidity)**
  On an extreme heat day with very high minimum and maximum temperatures and low humidity, the model predicts very high evaporation (~14.9 mm), with a 95% interval from about 10–20 mm. This is the most critical scenario for reservoir management.

- **6 July 2020 (mid-winter, cool and humid)**
  In mid-winter, with low temperatures and high humidity, predicted evaporation is very low (~2.3 mm), and the 95% interval ranges from near zero to about 6–7 mm.

These results align with physical intuition; evaporation is greatest on hot, dry summer days and smallest on cool, humid winter days.

Comparison Across Days and Operational Implications

Comparing the four scenarios:

- **Seasonal effect**: Even with similar humidity, summer days (especially January) produce substantially higher evaporation than winter days (July), reflecting stronger solar radiation and higher temperatures.

- **Temperature effect**: The difference between the December and January scenarios highlights the role of minimum temperature: very warm nights (as in January) contribute to significantly higher total daily evaporation.

- **Humidity effect**: For the February day with high humidity, predicted evaporation is notably lower than for the December or January days, despite being in summer, showing that humidity can partially offset the effect of season and temperature.

From an operational perspective, these results show that:

- Very hot, low-humidity summer days (like 13 January) are the key risk days when evaporation can exceed 10 mm by a large margin.
- Mild or humid summer days (like 29 February) can still have moderate evaporation but are less critical.
- Winter days pose little evaporation risk, even with occasional clear and dry conditions.

Evaporation Above 10 mm and Temporary Measures

MWC takes temporary measures when daily evaporation at Cardinia exceeds 10 mm. Using the 95% prediction intervals in Table 1:

- **Days for which we can say with 95% confidence that evaporation will exceed 10 mm**
    - 13 January 2020: the entire 95% prediction interval (≈ 10.1–19.6 mm) lies above 10 mm.
    - Conclusion: on this day, it is very likely that temporary transfer measures from Silvan Reservoir will be needed.
- **Days for which we can say with 95% confidence that evaporation will not exceed 10 mm**
    - 29 February 2020: the upper bound of the 95% interval is below 10 mm (~9.9 mm).
    - 6 July 2020: the upper bound is also well below 10 mm (~6.6 mm).
    - Conclusion: on these two days, temporary measures are very unlikely to be needed due to evaporation alone.

- **Day with uncertainty about exceeding 10 mm**
    - 25 December 2020: the 95% interval (≈ 4.2–13.0 mm) crosses the 10 mm threshold.
    - Conclusion: for this day the model is inconclusive; evaporation might be below 10 mm or above it. Operational planning would need to consider additional information or contingency arrangements.

CONCLUSION

This analysis has developed and justified a multiple linear regression model for daily evaporation at Melbourne's Cardinia Reservoir, based on a full financial year of historical weather data.

The key findings are:

- Evaporation is strongly seasonal, with highest values in summer (January–March) and lowest values in winter (June–August).
- Minimum temperature and 9 am relative humidity are the most important meteorological predictors: higher minimum temperatures increase evaporation, while higher humidity decreases it.
- The effect of humidity varies by month, motivating the inclusion of a Month × humidity interaction in the final model.
- Day of week and maximum temperature do not contribute meaningfully once month, minimum temperature, and humidity are accounted for.
- The final model explains around 64% of the variability in daily evaporation and satisfies the main regression assumptions to a reasonable degree for practical forecasting.

When applied to specific future weather scenarios, the model provides 95% prediction intervals for daily evaporation. These intervals allow MWC to identify days when evaporation is almost certainly below operational thresholds, days when it is almost certainly above, and days where the outcome is uncertain and may require closer monitoring.

In particular, very hot, low-humidity summer days such as 13 January 2020 are almost certain to produce evaporation above 10 mm, requiring temporary transfer measures from Silvan Reservoir. Mild or humid days, and winter days, are unlikely to reach this threshold.

The model can be incorporated into MWC's operational planning to support daily decisions on water transfers and storage management. Future work could extend the model to multiple years of data, include additional meteorological variables such as wind speed and solar radiation, or explore non-linear modelling approaches if stronger deviations from linearity are observed.

APPENDIX: R CODE

The following R code produces all data processing, modelling, diagnostics, and prediction steps described in the report.

```
# Melbourne Evaporation Case Study - R Code


# Load packages
library(tidyverse)


# 1. Data import and preparation


# Read data
melb <- read_csv("melbourne.csv")


# Clean and transform variables
```

```
melb <- melb %>%
  mutate(
    # Parse date
    Date  = as.Date(Date),
    # Derive Month and Day as ordered factors
    Month = factor(
      month.name[as.integer(format(Date, "%m"))],
      levels = month.name
    ),
    Day   = factor(
      weekdays(Date),
      levels = c("Monday","Tuesday","Wednesday","Thursday",
                 "Friday","Saturday","Sunday")
    )
  ) %>%

  # Rename key variables for convenience
  rename(
    evap    = `Evaporation (mm)`,
    mintemp = `Minimum temperature (Deg C)`,
    maxtemp = `Maximum Temperature (Deg C)`,
    rh9     = `9am relative humidity (%)`
  )

# Remove rows with missing evaporation
melb <- melb %>% filter(!is.na(evap))




# 2. Bivariate summaries

# Evaporation vs Month (boxplot)
ggplot(melb, aes(x = Month, y = evap)) +
```

```r
  geom_boxplot() +
  labs(
    title = "Daily evaporation by month",
    x = "Month",
    y = "Evaporation (mm)"
  )

# Evaporation vs Day of week (boxplot)
ggplot(melb, aes(x = Day, y = evap)) +
  geom_boxplot() +
  labs(
    title = "Daily evaporation by day of week",
    x = "Day of week",
    y = "Evaporation (mm)"
  )

# Evaporation vs minimum temperature (scatter with smooth)
ggplot(melb, aes(x = mintemp, y = evap)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Evaporation vs minimum temperature",
    x = "Minimum temperature (°C)",
    y = "Evaporation (mm)"
  )

# Evaporation vs maximum temperature (scatter with smooth)
ggplot(melb, aes(x = maxtemp, y = evap)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Evaporation vs maximum temperature",
    x = "Maximum temperature (°C)",
    y = "Evaporation (mm)"
```

```
  )

# Evaporation vs 9am relative humidity (scatter with smooth)
ggplot(melb, aes(x = rh9, y = evap)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Evaporation vs 9am relative humidity",
    x = "9am relative humidity (%)",
    y = "Evaporation (mm)"
  )

# 3. Model building and selection

# FULL model including ALL predictors and the interaction
mod0 <- lm(evap ~ Month + Day + mintemp + maxtemp + rh9 + Month:rh9,
           data = melb)

# Look at the output to see p-values of numeric predictors
summary(mod0)


#Test whether 'Day of the week' is significant
# Model without Day (for comparison)
mod_no_day <- lm(evap ~ Month + mintemp + maxtemp + rh9 + Month:rh9,
                 data = melb)

# F-test comparing mod_no_day vs full model
anova(mod_no_day, mod0)   # If p > 0.05, Day is NOT significant


# Since Day is not significant, we drop it and continue from mod_no_day
mod1 <- mod_no_day
```

```
#Test if the Month × humidity interaction is significant
# Model without interaction
mod_no_int <- lm(evap ~ Month + mintemp + maxtemp + rh9, data = melb)

# Compare models: If p < 0.05, interaction is significant and must stay
anova(mod_no_int, mod1)




# Test maximum temperature
# Model without maximum temperature
mod_no_max <- lm(evap ~ Month + mintemp + rh9 + Month:rh9,
                 data = melb)

# Compare models: If p > 0.05, maxtemp is NOT significant and can be
removed
anova(mod_no_max, mod1)




# FINAL model after removing Day and maxtemp
mod_final <- mod_no_max
summary(mod_final)




# 4. Model diagnostics

# Standard diagnostic plots
par(mfrow = c(2, 2))
plot(mod_final)
par(mfrow = c(1, 1))
```

```r
# 5. Predictions for specified days

# Construct new data for prediction
new_days <- tibble(
  Date    = as.Date(c("2020-02-29",
                      "2020-12-25",
                      "2020-01-13",
                      "2020-07-06")),
  mintemp = c(13.8, 16.4, 26.5, 6.8),
  maxtemp = c(23.2, 31.9, 44.3, 10.6),
  rh9     = c(74,   57,   35,   76)
) %>%
  mutate(
    Month = factor(
      month.name[as.integer(format(Date, "%m"))],
      levels = month.name
    )
  )

# Use final model to make predictions with 95% prediction intervals
predictions <- predict(
  mod_final,
  newdata = new_days,
  interval = "prediction",
  level = 0.95
)

# Combine with input data for a nice table
pred_table <- bind_cols(new_days, as_tibble(predictions))
pred_table

# round
pred_table_rounded <- pred_table %>%
  mutate(across(c(fit, lwr, upr), ~ round(.x, 2)))
```

```
pred_table_rounded





# End of R code
```

REFRENCES

- Bureau of Meteorology (BOM) 2024, Australian Government, viewed 27 November 2025, http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml.
- R Core Team 2024, R Foundation for Statistical Computing, Vienna, viewed 27 November 2025, https://www.r-project.org/.
- Wickham, H, Averick, M, Bryan, J, Chang, W, McGowan, L, François, R, Grolemund, G, et al. 2019, Journal of Open Source Software, viewed 27 November 2025, https://www.tidyverse.org.
- Australian Bureau of Statistics (ABS) 2023, Australian Government, viewed 27 November 2025, https://www.abs.gov.au/statistics/environment.
- United States Environmental Protection Agency (EPA) 2023, US Government, viewed 27 November 2025, https://www.epa.gov/water-research.
- University of Adelaide 2025, viewed 27 November 2025, https://myuni.adelaide.edu.au/courses/105626/modules.