

Informe de control de calidad de datos climáticos - Parcela 3

Ing. Miguel Silva

2025-12-06

Table of contents

1	Introducción	1
2	Descripción general de la base de datos	2
3	Análisis exploratorio	3
4	Control de calidad	3
5	Resultados	3
6	Visualización gráfica de las variables	7

1 Introducción

En este informe se presenta el control de calidad de la base de datos climáticos registrada en la Parcela 3, a partir del archivo `Datos_Parcels_3.xlsx`.

El objetivo principal es identificar y cuantificar:

- Datos faltantes por variable.
- Valores fuera de rangos físicos razonables.
- Registros atípicos que puedan comprometer análisis posteriores (balance de energía, estimación de ET, etc.).
- Visualizar el comportamiento temporal de las variables.

2 Descripción general de la base de datos

La base de datos esta conformada por los registros horarios de las siguientes variables:

Variables meteorológicas:

- Temperatura (°C)
- Humedad relativa (%)
- Velocidad del viento (m/s)
- Dirección del viento (°)
- Presión atmosférica (mbar)
- Radiación global (Wm-2)
- Precipitación (mm)

Variables de flujos de energía:

- Temperatura superficial (°C)
- Flujo de energía en el suelo (Wm-2)
- Radiación Neta (Wm-2)
- Temperatura termocupla 1 (°C)
- Temperatura termocupla 2 (°C)
- Humedad del suelo a los 5 cm (m3/m3)
- Humedad del suelo a los 40 cm (m3/m3)

Para su procesamiento preliminar se aplica lo siguiente:

```
# A tibble: 1 x 3
  fecha_inicio      fecha_fin      n_registros
  <dtm>            <dtm>            <int>
1 2025-07-21 14:00:00 2025-11-28 15:00:00      2438

# A tibble: 1 x 9
  n_Temperatura_aire_C n_Humedad_relativa_porc n_Precipitacion_mm
  <int>                <int>                <int>
1      2434            2434            2434
# i 6 more variables: n_Velocidad_viento_ms <int>, n_Radiacion_neta_Wm2 <int>,
#   n_Radiacion_global_Wm2 <int>, n_Temperatura_superficie_C <int>,
#   n_Humedad_suelo_5cm_m3m3 <int>, n_Humedad_suelo_40cm_m3m3 <int>
```

3 Análisis exploratorio

Este análisis primero identifica el período temporal cubierto, el número de registros y variables, y determina qué columnas contienen información numérica. Luego calcula estadísticas descriptivas básicas; media, mediana, desviación estándar, mínimos, máximos y cantidad de datos faltantes para todas las variables, permitiendo detectar rangos típicos, dispersión y posibles anomalías. Finalmente, cuantifica los valores ausentes por variable, lo cual ayuda a evaluar la completitud del conjunto de datos antes de aplicar controles de calidad más estrictos.

4 Control de calidad

Las reglas de control de calidad aplicadas en este informe son:

Continuidad temporal

Verificación de que la serie está ordenada y sin duplicados en fecha_hora.

Rangos físicos razonables (pueden ajustarse a tus criterios locales):

Temperatura_aire_C: -10 a 45 °C

Humedad_relativa_porc: 0 a 100 %

Precipitacion_mm: 0 mm

Velocidad_viento_ms: 0 m/s

Radiacion_neta_Wm2: -200 a 1000 W/m²

Tratamiento de datos

Los valores fuera de rango se marcan mediante flags y, en la base validada, se reemplazan por NA para evitar su uso directo en análisis.

5 Resultados

Resultados del análisis exploratorio

```
resultados_ae <- analisis_exploratorio(datos_brutos)

#Info básica en tabla

info_basica <- as.data.frame(resultados_ae$info_basica)
info_basica

  n_filas n_columnas fecha_inicio fecha_fin duracion_dias
1    2438         32   1753106400 1764342000      130.0417

#| label: estadisticas-principales
#| echo: true

estadisticas_df <- purrr::map_dfr(
  resultados_ae$estadisticas,
  ~ as.data.frame(.x),
```

```
.id = "variable"
)
```

```
estadisticas_df
```

	variable	media	mediana	desviacion	min
1	Radiacion_global_Wm2	1.891088e+02	10.8350	2.617007e+02	0.000
2	Precipitacion_mm	2.940427e-02	0.0000	2.730409e-01	0.000
3	Rayos_conteo	2.875924e-03	0.0000	8.354031e-02	0.000
4	Velocidad_viento_ms	5.160140e-01	0.4580	3.449411e-01	0.048
5	Direccion_viento_grados	1.761297e+02	163.9000	7.951649e+01	0.900
6	Rachas_ms	1.511889e+00	1.3470	1.043812e+00	0.123
7	Temperatura_aire_C	1.559380e+01	14.2250	6.835956e+00	2.317
8	Presion_atmosferica_mbar	1.016620e+03	1016.3310	3.036572e+00	1009.648
9	Humedad_relativa_porc	6.761639e+01	70.6500	2.316700e+01	15.700
10	Radiacion_global_MJm2	6.816402e-01	0.0366	9.424293e-01	0.000
11	Temperatura_superficie_C	1.596208e+01	13.7500	8.175708e+00	1.126
12	Flujo_energia_suelo_Wm2	-9.835949e-02	-1.0150	4.760450e+00	-10.820
13	Radiacion_neta_Wm2	5.264392e+02	317.9000	7.687920e+02	-1642.000
14	Temperatura_termocupla1_C	1.658387e+01	14.1500	8.853681e+00	0.719
15	Temperatura_termocupla2_C	1.658374e+01	14.1500	8.853524e+00	0.721
16	Humedad_suelo_5cm_m3m3	3.570538e-01	0.3650	3.624735e-02	0.214
17	Humedad_suelo_40cm_m3m3	4.897650e-01	0.4910	5.753877e-03	0.476

```
max nas
```

1	873.0000	0
2	6.0350	4
3	3.0000	4
4	2.1030	4
5	359.7000	4
6	6.3280	4
7	32.9200	4
8	1027.0980	4
9	100.0000	4
10	3.1428	4
11	36.3800	4
12	12.4600	4
13	3004.0000	4
14	38.4300	4
15	38.4300	4
16	0.4500	4
17	0.5010	4

```
#| label: valores-faltantes
```

```
#| echo: true
```

```
valores_faltantes_df <- data.frame(
  variable = names(resultados_ae$valores_faltantes),
  n_na = as.numeric(resultados_ae$valores_faltantes)
)
```

```
valores_faltantes_df
```

	variable	n_na
1	fecha_hora	105
2	Precipitacion_mm	4
3	Rayos_conteo	4
4	Dist_km_Avg	4
5	Velocidad_viento_ms	4
6	Direccion_viento_grados	4
7	Rachas_ms	4
8	Temperatura_aire_C	4
9	VP_mbar_Avg	4
10	Presion_atmosferica_mbar	4
11	ETos	4
12	Rso	4
13	Humedad_relativa_porc	4
14	RHT_C	4
15	TiltNS_deg_Avg	4
16	TiltWE_deg_Avg	4
17	Radiacion_global_MJm2	4
18	CVMeta	14
19	Invalid_Wind	4
20	TT_C_Avg	4
21	Temperatura_superficie_C	4
22	Flujo_energia_suelo_Wm2	4
23	Radiacion_neta_Wm2	4
24	CNR_Wm2_Avg	4
25	Temperatura_termocupla1_C	4
26	Temperatura_termocupla2_C	4
27	Humedad_suelo_5cm_m3m3	4
28	Humedad_suelo_40cm_m3m3	4

Resumen de flags por variable

```
# A tibble: 1 x 15
  temp_ok temp_fuera temp_na hr_ok hr_fuera hr_na pp_ok pp_negativa pp_na ws_ok
  <int>      <int>    <int> <int>    <int> <int> <int>      <int> <int> <int>
1   2434         0      4  2434         0    4  2434         0    4  2434
# i 5 more variables: ws_negativa <int>, ws_na <int>, rn_ok <int>,
#   rn_fuera <int>, rn_na <int>

# A tibble: 1 x 15
  temp_ok temp_fuera temp_na hr_ok hr_fuera hr_na pp_ok pp_negativa pp_na ws_ok
  <dbl>      <dbl>    <dbl> <dbl>    <dbl> <dbl> <dbl>      <dbl> <dbl> <dbl>
1   99.8         0  0.164  99.8         0 0.164  99.8         0 0.164  99.8
# i 5 more variables: ws_negativa <dbl>, ws_na <dbl>, rn_ok <dbl>,
#   rn_fuera <dbl>, rn_na <dbl>
```

Resultados del control de calidad

```
resultados_cc <- control_calidad(datos_brutos)
```

```

fuera_rango_df <- purrr::map_dfr(
  resultados_cc$fuera_rango,
  ~ as.data.frame(.x),
  .id = "variable"
)

atipicos_iqr_df <- purrr::map_dfr(
  resultados_cc$atipicos_iqr,
  ~ as.data.frame(.x),
  .id = "variable"
)

completitud_df <- data.frame(
  variable = names(resultados_cc$completitud),
  completitud = as.numeric(resultados_cc$completitud)
)

list(
  fuera_rango = fuera_rango_df,
  atipicos_iqr = atipicos_iqr_df,
  completitud = completitud_df,
  saltos_tiempo = resultados_cc$saltos_temporales
)

```

\$fuera_rango

	variable	n_fuera_rango	porcentaje
1	Temperatura_aire_C	0	0
2	Humedad_relativa_porc	0	0
3	Velocidad_viento_ms	0	0
4	Radiacion_global_Wm2	0	0
5	Precipitacion_mm	0	0
6	Humedad_suelo_5cm_m3m3	0	0
7	Humedad_suelo_40cm_m3m3	0	0

\$atipicos_iqr

	variable	n_atipicos	porcentaje
1	Radiacion_global_Wm2	0	0.00
2	Precipitacion_mm	179	7.34
3	Rayos_conteo	3	0.12
4	Velocidad_viento_ms	10	0.41
5	Direccion_viento_grados	0	0.00
6	Rachas_ms	4	0.16
7	Temperatura_aire_C	0	0.00
8	Presion_atmosferica_mbar	28	1.15
9	Humedad_relativa_porc	0	0.00
10	Radiacion_global_MJm2	0	0.00
11	Temperatura_superficie_C	0	0.00
12	Flujo_energia_suelo_Wm2	0	0.00
13	Radiacion_neta_Wm2	66	2.71
14	Temperatura_termocupla1_C	0	0.00
15	Temperatura_termocupla2_C	0	0.00

16	Humedad_suelo_5cm_m3m3	190	7.79
17	Humedad_suelo_40cm_m3m3	0	0.00

\$completitud

	variable	completitud
1	Radiacion_global_Wm2	100.00
2	Precipitacion_mm	99.84
3	Rayos_conteo	99.84
4	Velocidad_viento_ms	99.84
5	Direccion_viento_grados	99.84
6	Rachas_ms	99.84
7	Temperatura_aire_C	99.84
8	Presion_atmosferica_mbar	99.84
9	Humedad_relativa_porc	99.84
10	Radiacion_global_MJm2	99.84
11	Temperatura_superficie_C	99.84
12	Flujo_energia_suelo_Wm2	99.84
13	Radiacion_neta_Wm2	99.84
14	Temperatura_termocupla1_C	99.84
15	Temperatura_termocupla2_C	99.84
16	Humedad_suelo_5cm_m3m3	99.84
17	Humedad_suelo_40cm_m3m3	99.84

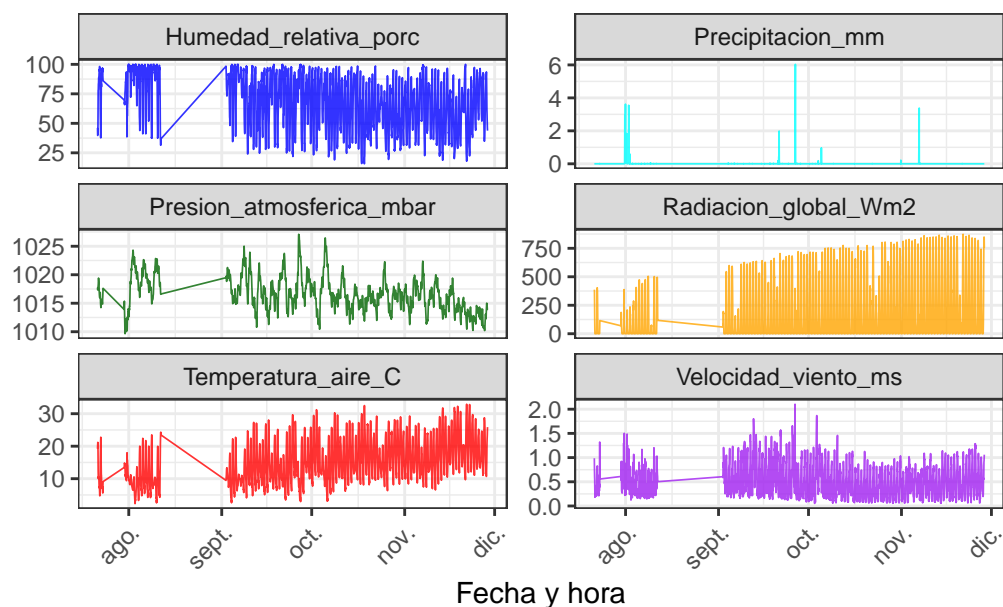
\$saltos_tiempo

[1] 103

6 Visualización gráfica de las variables

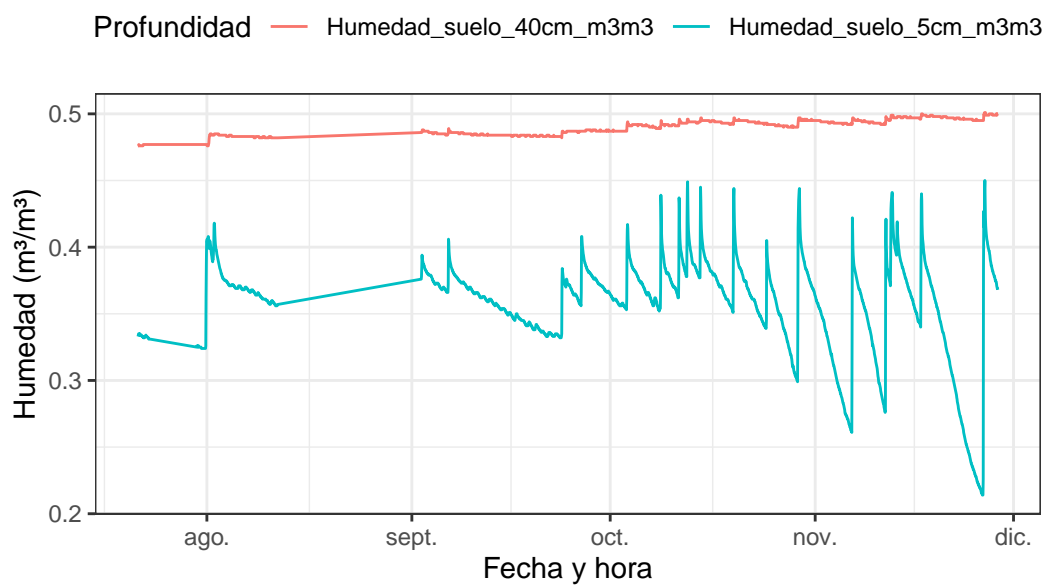
Series temporales facetadas de variables principales

Series temporales de variables meteorológicas – Parcela 3



Humedad del suelo en el tiempo (2 profundidades)

Humedad del suelo a diferentes profundidades



Precipitación diaria acumulada

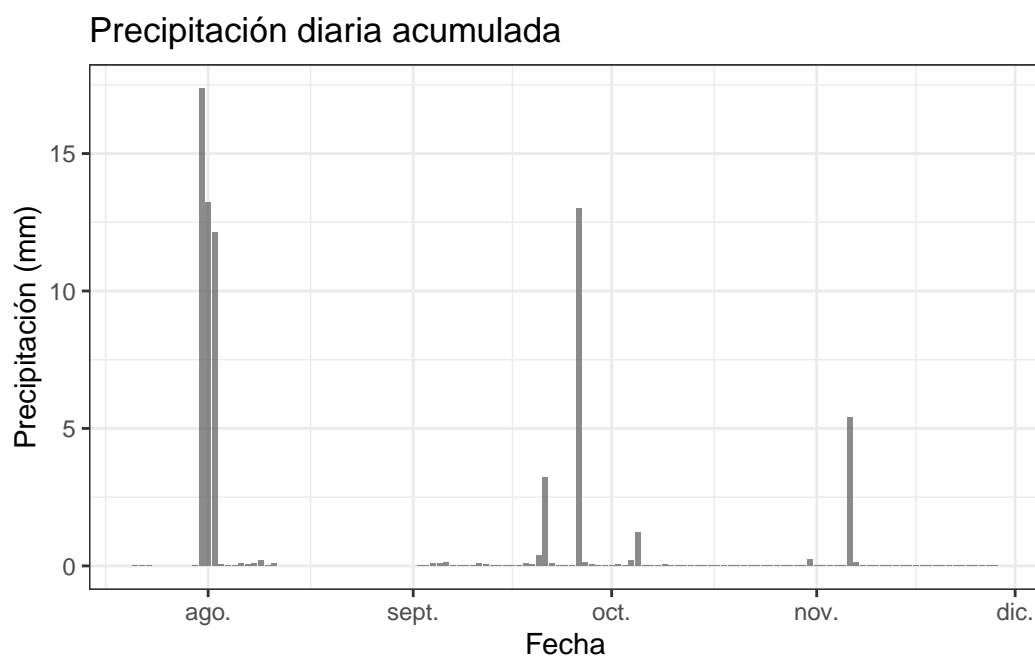
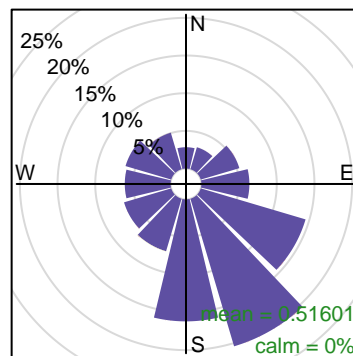


Figure 1: Precipitación diaria acumulada.

Rosa de vientos

Rosa de vientos – Parcela 3



0 to 22 to 44 to 6
Velocidad (m s^{-1})
(m s^{-1})

Frequency of counts by wind direction (%)

Figure 2: Rosa de vientos (dirección y velocidad del viento).

Distribución de temperaturas

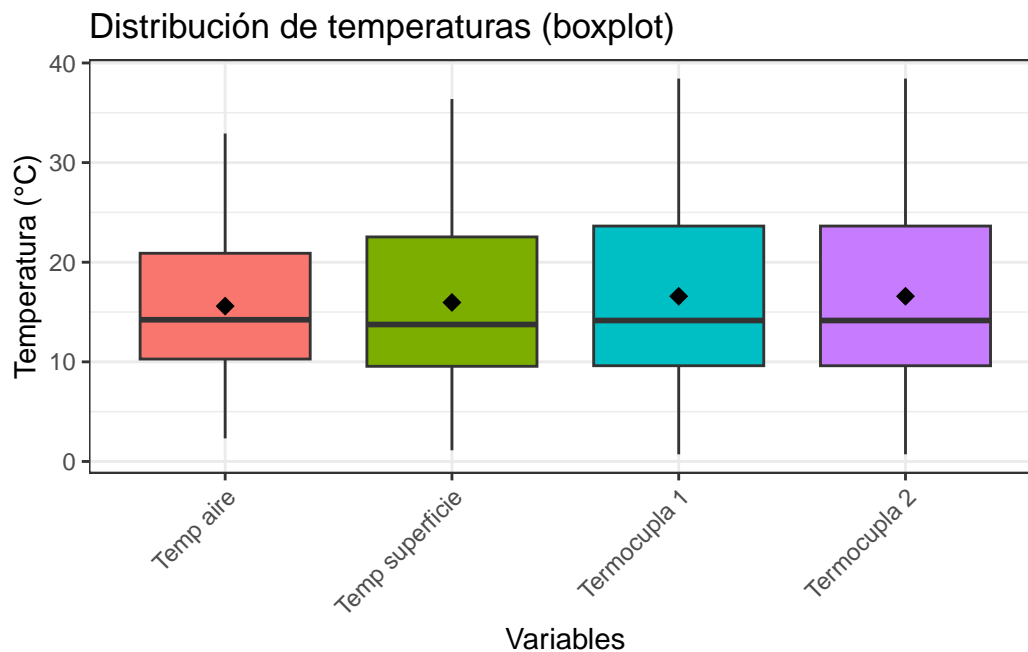


Figure 3: Distribución de temperaturas del aire, superficie y termocuplas.

Generación de base de datos validada

```
datos_validados <- datos_qc %>%
mutate(
  Temperatura_aire_C = if_else(flag_temp_rango == "ok", Temperatura_aire_C, NA_real_),
  Humedad_relativa_porc = if_else(flag_hr_rango == "ok", Humedad_relativa_porc, NA_real_),
```

```

Precipitacion_mm = if_else(flag_pp_rango == "ok", Precipitacion_mm, NA_real_),
Velocidad_viento_ms = if_else(flag_ws_rango == "ok", Velocidad_viento_ms, NA_real_),
Radiacion_neta_Wm2 = if_else(flag_rn_rango == "ok", Radiacion_neta_Wm2, NA_real_)
)

readr::write_csv(datos_validados, "Datos_Parcela_3_validado.csv")

datos_validados %>%
summarise(
n_temp_validas = sum(!is.na(Temperatura_aire_C)),
n_hr_validas = sum(!is.na(Humedad_relativa_porc)),
n_pp_validas = sum(!is.na(Precipitacion_mm)),
n_ws_validas = sum(!is.na(Velocidad_viento_ms)),
n_rn_validas = sum(!is.na(Radiacion_neta_Wm2))
)

```

A tibble: 1 x 5

	n_temp_validas	n_hr_validas	n_pp_validas	n_ws_validas	n_rn_validas
	<int>	<int>	<int>	<int>	<int>
1	2434	2434	2434	2434	1821

{#}