

Time Window

October 30, 2025

1 Time Window

Jakob Balkovec

Finding the right time window for temporal and spatial dataset.

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[2]: TEMPORAL = r'/Users/jbalkovec/Desktop/MDR/Temporal/Pipeline/data/master/
↳final_master.csv'
SPATIAL = r'/Users/jbalkovec/Desktop/MDR/Spatial/processed/
↳ground_cleaned_lind_2021.csv'
```

```
[14]: temporal_df = pd.read_csv(TEMPORAL)
spatial_df = pd.read_csv(SPATIAL)

temporal_df["date"] = pd.to_datetime(temporal_df["date"]).dt.tz_localize(None)
spatial_df["date"] = pd.to_datetime(spatial_df["date"]).dt.tz_localize(None).dt.
↳floor("D")
```

```
[15]: temporal_df.head()
```

```
[15]:
```

	date	station_id	crx_vn	longitude	latitude	air_temp_max	\
0	2007-07-31	4136	1.302	-117.53	47.42	NaN	
1	2007-08-01	4136	1.302	-117.53	47.42	32.9	
2	2007-08-02	4136	1.302	-117.53	47.42	35.4	
3	2007-08-03	4136	1.302	-117.53	47.42	29.4	
4	2007-08-04	4136	1.302	-117.53	47.42	27.6	

	air_temp_min	air_temp_mean	air_temp_avg	precipitation	...	\
0	NaN	NaN	NaN	NaN	...	
1	2.4	17.7	19.4	0.0	...	
2	4.0	19.7	22.0	0.0	...	
3	7.5	18.4	21.5	0.0	...	
4	6.5	17.0	18.0	0.0	...	

	soil_temp_50cm	soil_temp_100cm	source_file	\
0	NaN	NaN	uscrn_WA_Spokane_17_SSW_2007.txt	
1	NaN	NaN	uscrn_WA_Spokane_17_SSW_2007.txt	
2	NaN	NaN	uscrn_WA_Spokane_17_SSW_2007.txt	
3	NaN	NaN	uscrn_WA_Spokane_17_SSW_2007.txt	
4	NaN	NaN	uscrn_WA_Spokane_17_SSW_2007.txt	

	LST	NDVI	Rain_sat	DOY	Rain_3d	SM_prev	SM_label
0	310.438308	0.495671	0.000057	212	0.0	NaN	NaN
1	307.114121	0.470536	0.016377	213	0.0	NaN	NaN
2	307.114121	0.470536	0.016377	214	0.0	NaN	NaN
3	307.114121	0.470536	0.016377	215	0.0	NaN	NaN
4	307.114121	0.470536	0.016377	216	0.0	NaN	NaN

[5 rows x 36 columns]

```
[16]: spatial_df.head()
```

```
[16]:
```

	date	station_id	station_name	precip_in	air_temp_obs_f	\
0	2025-09-25	2021	Lind #1	0.0	79.3	
1	2025-09-25	2021	Lind #1	0.0	81.3	
2	2025-09-25	2021	Lind #1	0.0	81.0	
3	2025-09-25	2021	Lind #1	0.0	79.9	
4	2025-09-25	2021	Lind #1	0.0	74.3	

	air_temp_avg_f	air_temp_max_f	air_temp_min_f	soil_moisture_2in	\
0	79.9		81.3	79.3	0.0
1	81.7		82.4	81.0	0.0
2	81.5		81.9	79.9	0.0
3	78.6		79.9	74.3	0.0
4	72.0		74.3	66.2	0.0

	soil_moisture_4in	...	air_temp_obs_c	air_temp_avg_c	air_temp_max_c	\
0	0.1	...	26.277778	26.611111	27.388889	
1	0.0	...	27.388889	27.611111	28.000000	
2	0.2	...	27.222222	27.500000	27.722222	
3	0.2	...	26.611111	25.888889	26.611111	
4	0.2	...	23.500000	22.222222	23.500000	

	air_temp_min_c	latitude	longitude	elevation_ft	state	network	county
0	26.277778	47.0	-118.57	1650	WA	SCAN	Adams
1	27.222222	47.0	-118.57	1650	WA	SCAN	Adams
2	26.611111	47.0	-118.57	1650	WA	SCAN	Adams
3	23.500000	47.0	-118.57	1650	WA	SCAN	Adams
4	19.000000	47.0	-118.57	1650	WA	SCAN	Adams

[5 rows x 35 columns]

```
[17]: # Span
print("Temporal:", temporal_df["date"].min(), "to", temporal_df["date"].max())
print("Spatial:", spatial_df["date"].min(), "to", spatial_df["date"].max())
```

Temporal: 2007-01-01 00:00:00 to 2025-10-23 00:00:00
 Spatial: 2025-09-25 00:00:00 to 2025-10-09 00:00:00

```
[18]: time_df = pd.merge(temporal_df, spatial_df, on='date', how='inner',
    ↪suffixes=('_temp', '_spat'))
```

```
[22]: print("Merged date range:", time_df["date"].min(), "to", time_df["date"].max())
print("Total overlapping rows:", len(time_df))
```

Merged date range: 2025-09-25 00:00:00 to 2025-10-09 00:00:00
 Total overlapping rows: 1005

1.1 Final Window (As of Oct 30th 2025)

Dataset	Start Date	End Date	Overlap Range
Temporal	2007-01-01	2025-10-23	—
Spatial	2025-09-25	2025-10-09	—
Overlap	2025-09-25	2025-10-09	Shared active window

In other words, the final overlapping time window is from September 25th, 2025 to October 9th, 2025.