Exploratory Data Analysis (EDA)

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, employing data visualization.

This notebook includes:

- EDA: missingness, distributions, feature–feature correlations, feature–target correlations
- Per-Country EDA (DE/FR)
- Time Series analysis (global & per-country): rolling means, autocorrelation function (ACF), cross-correlation with target
- Simple feature engineering: simple imputation comparison, simple selection, mutual information(MI), principal component analysis (PCA)

1. Configuration and Preparing

In [1]:
```
# ==== Configuration ====
DATA_DIR = "../data/"
X_TRAIN_PATH = f"{DATA_DIR}X_train_NHkHMNU.csv"
Y_TRAIN_PATH = f"{DATA_DIR}y_train_ZAN5mwg.csv"

# Time series analysis config
ROLL_WINDOWS = [7, 14, 30]

RANDOM_STATE = 47
```

In [2]:
```
# ==== Imports ====
import numpy as np
import pandas as pd
import math
import matplotlib.pyplot as plt
import warnings

from sklearn.impute import SimpleImputer, KNNImputer
from sklearn.preprocessing import StandardScaler
```

```
from sklearn.feature_selection import VarianceThreshold, mutual_info_regression
from sklearn.decomposition import PCA
from sklearn.model_selection import train_test_split
from sklearn.ensemble import HistGradientBoostingRegressor
import seaborn as sns

warnings.filterwarnings('ignore')
plt.rcParams['figure.figsize'] = (8, 5)  # one chart per figure
```

In [3]:
```
# ==== 1) Load data & basic checks ====
X = pd.read_csv(X_TRAIN_PATH)
y = pd.read_csv(Y_TRAIN_PATH)

assert "ID" in X.columns and "ID" in y.columns, "ID column must exist in both X and y"
df = X.merge(y, on="ID", how="left")

print("Shapes -> X:", X.shape, " y:", y.shape, " merged:", df.shape)
print("Columns:", list(df.columns))

display(df.head())
```

```
Shapes -> X: (1494, 35)  y: (1494, 2)  merged: (1494, 36)
Columns: ['ID', 'DAY_ID', 'COUNTRY', 'DE_CONSUMPTION', 'FR_CONSUMPTION', 'DE_FR_EXCHANGE', 'FR_DE_EXCHANGE', 'DE_NET_EXPORT', 'FR_NET_EXPORT', 'DE_NET_IMPORT', 'FR_NET_IMPORT', 'DE_GAS', 'FR_GAS', 'DE_COAL', 'FR_COAL', 'DE_HYDRO', 'FR_HYDRO', 'DE_NUCLEAR', 'FR_NUCLEAR', 'DE_SOLAR', 'FR_SOLAR', 'DE_WINDPOW', 'FR_WINDPOW', 'DE_LIGNITE', 'DE_RESIDUAL_LOAD', 'FR_RESIDUAL_LOAD', 'DE_RAIN', 'FR_RAIN', 'DE_WIND', 'FR_WIND', 'DE_TEMP', 'FR_TEMP', 'GAS_RET', 'COAL_RET', 'CARBON_RET', 'TARGET']
```

	ID	DAY_ID	COUNTRY	DE_CONSUMPTION	FR_CONSUMPTION	DE_FR_EXCHANGE	FR_DE_EXCHANGE	DE_NET_EXPORT	FR_NET_EXPOR
0	1054	206	FR	0.210099	-0.427458	-0.606523	0.606523	NaN	0.69286
1	2049	501	FR	-0.022399	-1.003452	-0.022063	0.022063	-0.573520	-1.13083
2	1924	687	FR	1.395035	1.978665	1.021305	-1.021305	-0.622021	-1.68258
3	297	720	DE	-0.983324	-0.849198	-0.839586	0.839586	-0.270870	0.56323
4	1101	818	FR	0.143807	-0.617038	-0.924990	0.924990	NaN	0.99032

5 rows × 36 columns

In [4]:
```
print("\nCOUNTRY value counts:")
if "COUNTRY" in df.columns:
    print(df["COUNTRY"].value_counts(dropna=False))
else:
    print("No COUNTRY column found.")

NON_FEATURE = {"ID", "COUNTRY", "TARGET"}
num_cols = [c for c in df.columns if c not in NON_FEATURE and pd.api.types.is_numeric_dtype(df[c])]
print("Numeric feature count:", len(num_cols))
```

```
COUNTRY value counts:
COUNTRY
FR    851
DE    643
Name: count, dtype: int64
Numeric feature count: 33
```

2. Missing Data Profiling

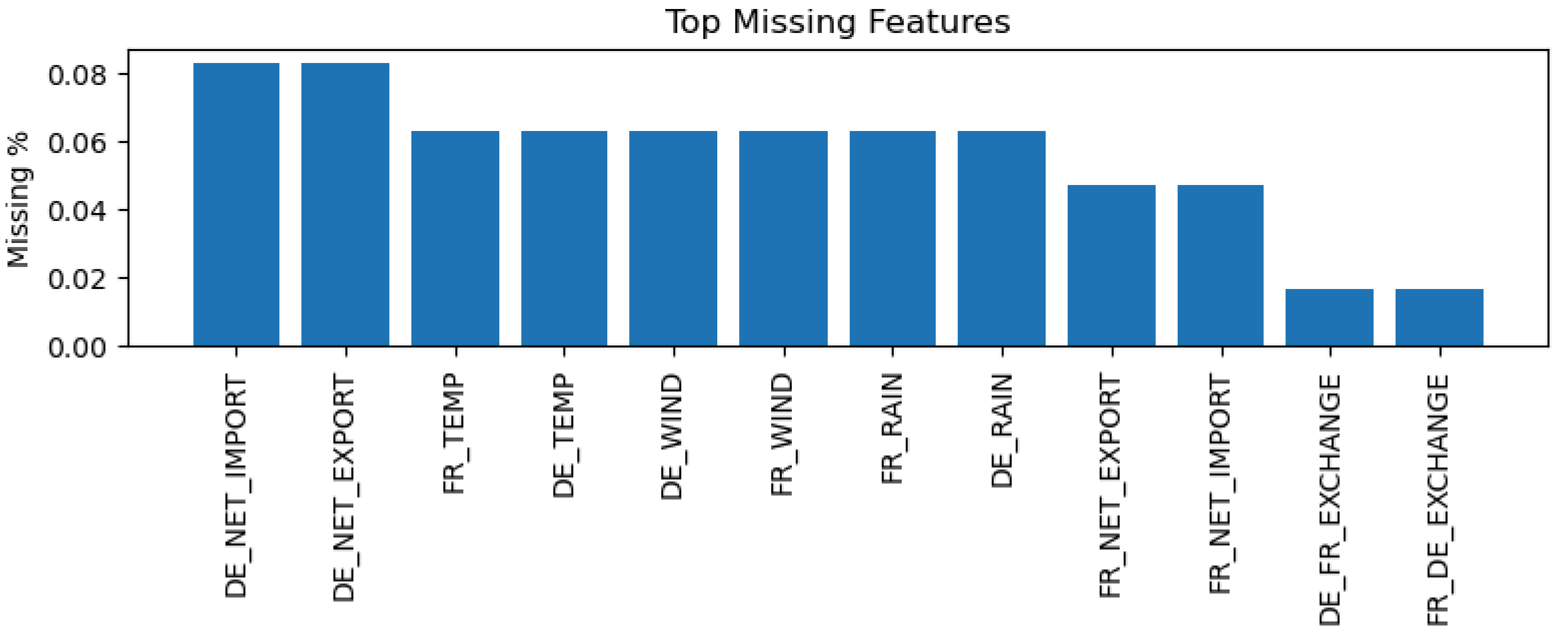
In [5]:
```
missing_counts = df.isna().sum().sort_values(ascending=False)
missing_pct = (missing_counts / len(df)).sort_values(ascending=False)
missing_df = pd.DataFrame({"missing_count": missing_counts, "missing_pct": missing_pct})
display(missing_df)
```

	missing_count	missing_pct
DE_NET_IMPORT	124	0.082999
DE_NET_EXPORT	124	0.082999
FR_TEMP	94	0.062918
DE_TEMP	94	0.062918
DE_WIND	94	0.062918
FR_WIND	94	0.062918
FR_RAIN	94	0.062918
DE_RAIN	94	0.062918
FR_NET_EXPORT	70	0.046854
FR_NET_IMPORT	70	0.046854
DE_FR_EXCHANGE	25	0.016734
FR_DE_EXCHANGE	25	0.016734
DE_CONSUMPTION	0	0.000000
COUNTRY	0	0.000000
DAY_ID	0	0.000000
ID	0	0.000000
FR_GAS	0	0.000000
DE_COAL	0	0.000000
FR_CONSUMPTION	0	0.000000
DE_GAS	0	0.000000
DE_SOLAR	0	0.000000
FR_NUCLEAR	0	0.000000

	missing_count	missing_pct
DE_NUCLEAR	0	0.000000
FR_HYDRO	0	0.000000
DE_HYDRO	0	0.000000
FR_COAL	0	0.000000
FR_SOLAR	0	0.000000
DE_WINDPOW	0	0.000000
DE_RESIDUAL_LOAD	0	0.000000
FR_RESIDUAL_LOAD	0	0.000000
FR_WINDPOW	0	0.000000
DE_LIGNITE	0	0.000000
GAS_RET	0	0.000000
COAL_RET	0	0.000000
CARBON_RET	0	0.000000
TARGET	0	0.000000

In [6]:

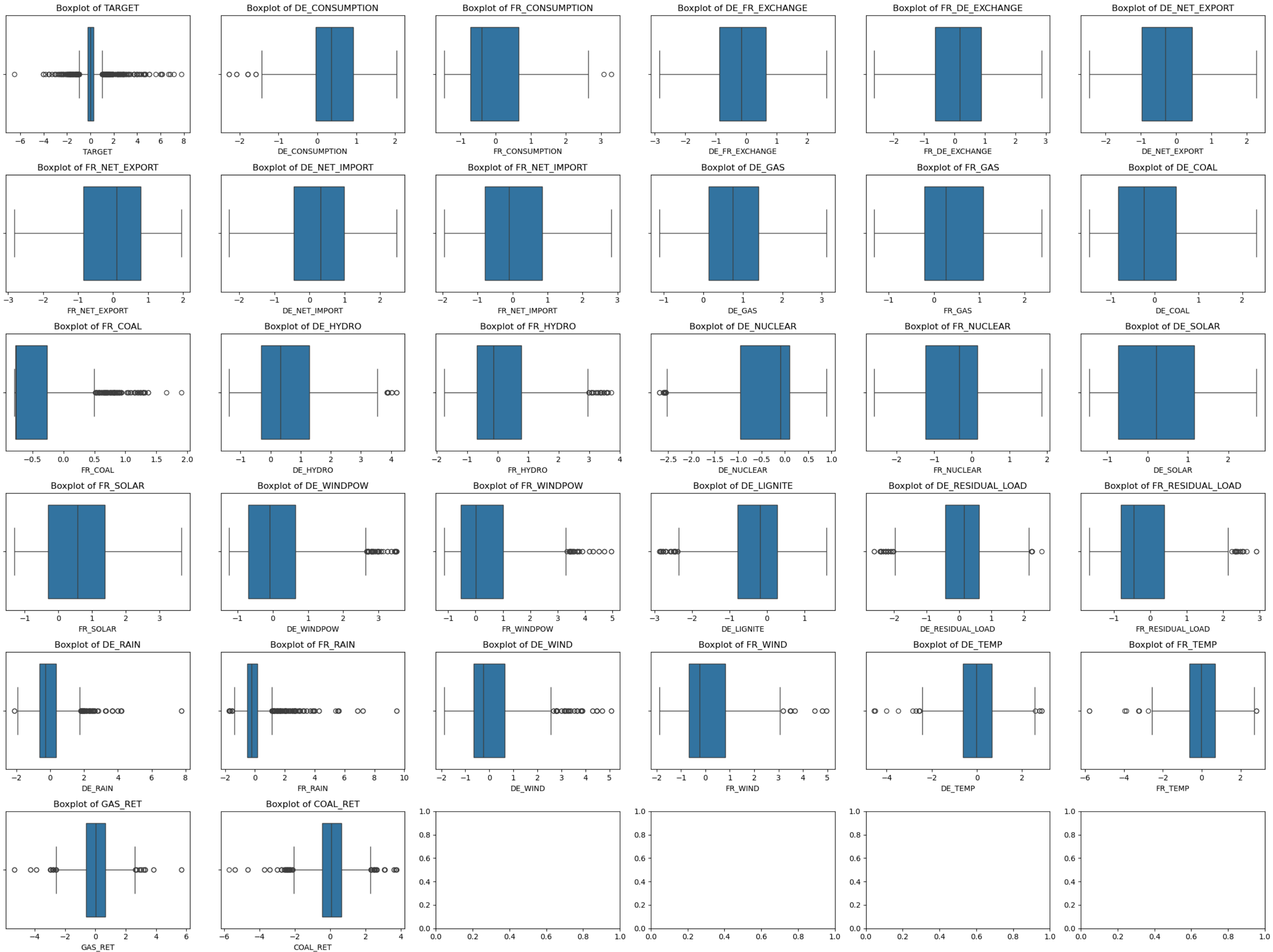
```
# Plot top missing features — single figure
top_miss = missing_df.head((missing_df['missing_count'] > 0).sum())
plt.figure()
plt.subplot(2, 1, 1)
plt.bar(top_miss.index.astype(str), top_miss["missing_pct"].values)
plt.xticks(rotation=90)
plt.ylabel("Missing %")
plt.title("Top Missing Features")
plt.tight_layout()
plt.show()
```



3. Check Outlier Data

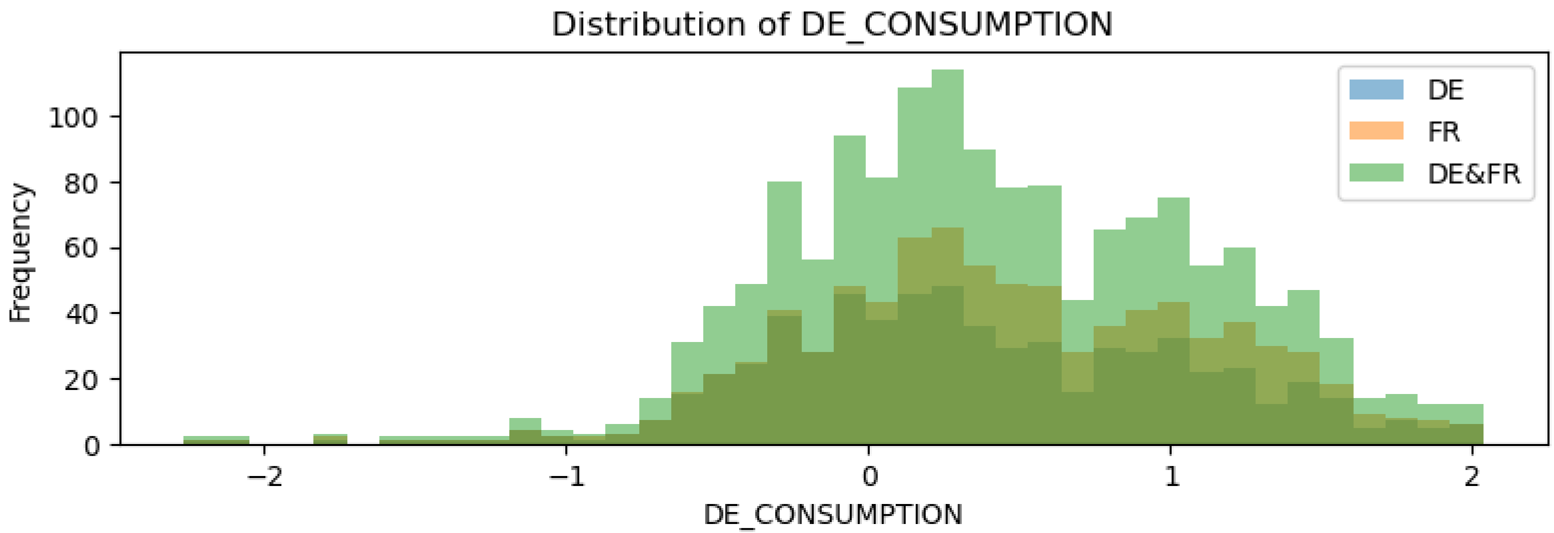
In [7]:

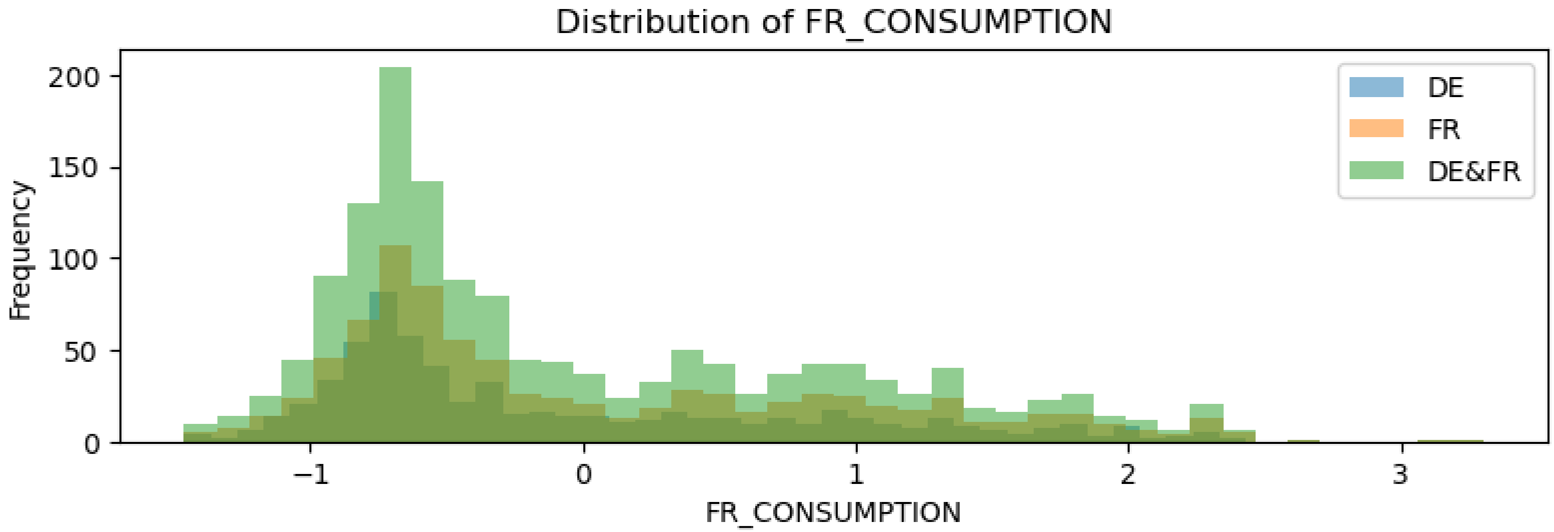
```
plots = [("TARGET", y["TARGET"])] + [(c, X[c]) for c in num_cols[1:-1]]

K = len(plots)
N = math.ceil(math.sqrt(K))
M = math.ceil(K / N)

fig, axes = plt.subplots(M, N, figsize=(N*4, M*3), squeeze=False)
axes = axes.ravel()

for i, (name, s) in enumerate(plots):
    sns.boxplot(x=s, ax=axes[i])
    axes[i].set_title(f"Boxplot of {name}")

plt.tight_layout()
plt.show()
```

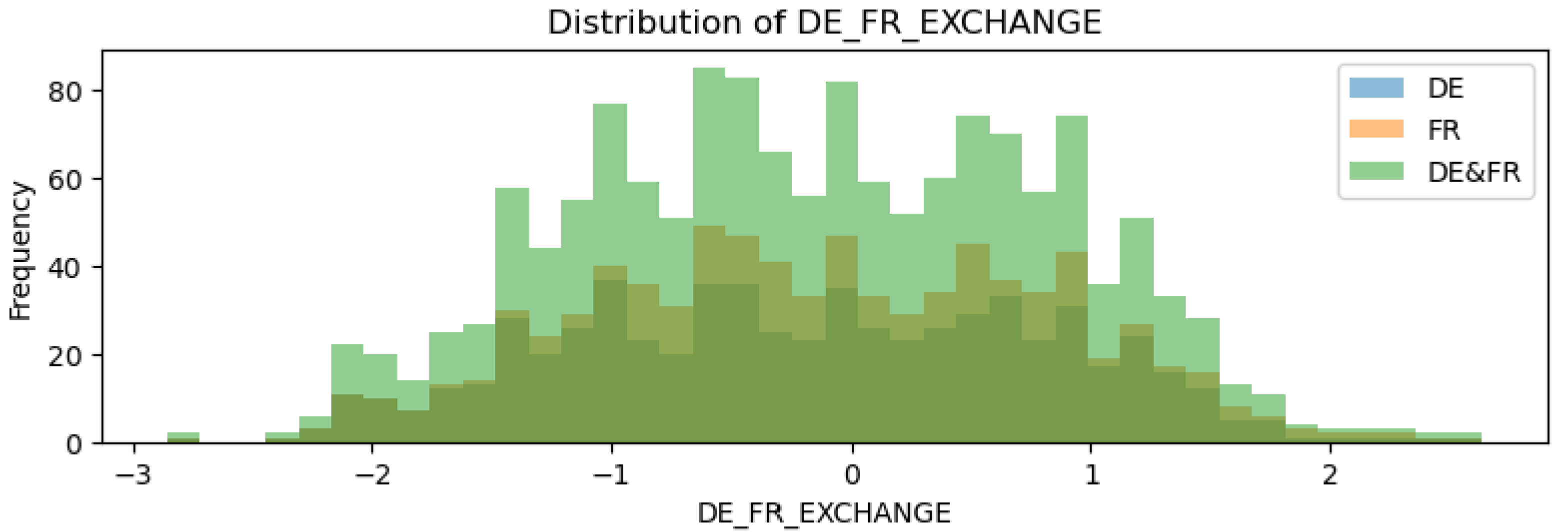

Boxplot of TARGET
TARGET
Boxplot of DE_CONSUMPTION
DE_CONSUMPTION
Boxplot of FR_CONSUMPTION
FR_CONSUMPTION
Boxplot of DE_FR_EXCHANGE
DE_FR_EXCHANGE
Boxplot of FR_DE_EXCHANGE
FR_DE_EXCHANGE
Boxplot of DE_NET_EXPORT
DE_NET_EXPORT
Boxplot of FR_NET_EXPORT
FR_NET_EXPORT
Boxplot of DE_NET_IMPORT
DE_NET_IMPORT
Boxplot of FR_NET_IMPORT
FR_NET_IMPORT
Boxplot of DE_GAS
DE_GAS
Boxplot of FR_GAS
FR_GAS
Boxplot of DE_COAL
DE_COAL
Boxplot of FR_COAL
FR_COAL
Boxplot of DE_HYDRO
DE_HYDRO
Boxplot of FR_HYDRO
FR_HYDRO
Boxplot of DE_NUCLEAR
DE_NUCLEAR
Boxplot of FR_NUCLEAR
FR_NUCLEAR
Boxplot of DE_SOLAR
DE_SOLAR
Boxplot of FR_SOLAR
FR_SOLAR
Boxplot of DE_WINDPOW
DE_WINDPOW
Boxplot of FR_WINDPOW
FR_WINDPOW
Boxplot of DE_LIGNITE
DE_LIGNITE
Boxplot of DE_RESIDUAL_LOAD
DE_RESIDUAL_LOAD
Boxplot of FR_RESIDUAL_LOAD
FR_RESIDUAL_LOAD
Boxplot of DE_RAIN
DE_RAIN
Boxplot of FR_RAIN
FR_RAIN
Boxplot of DE_WIND
DE_WIND
Boxplot of FR_WIND
FR_WIND
Boxplot of DE_TEMP
DE_TEMP
Boxplot of FR_TEMP
FR_TEMP
Boxplot of GAS_RET
GAS_RET
Boxplot of COAL_RET
COAL_RET

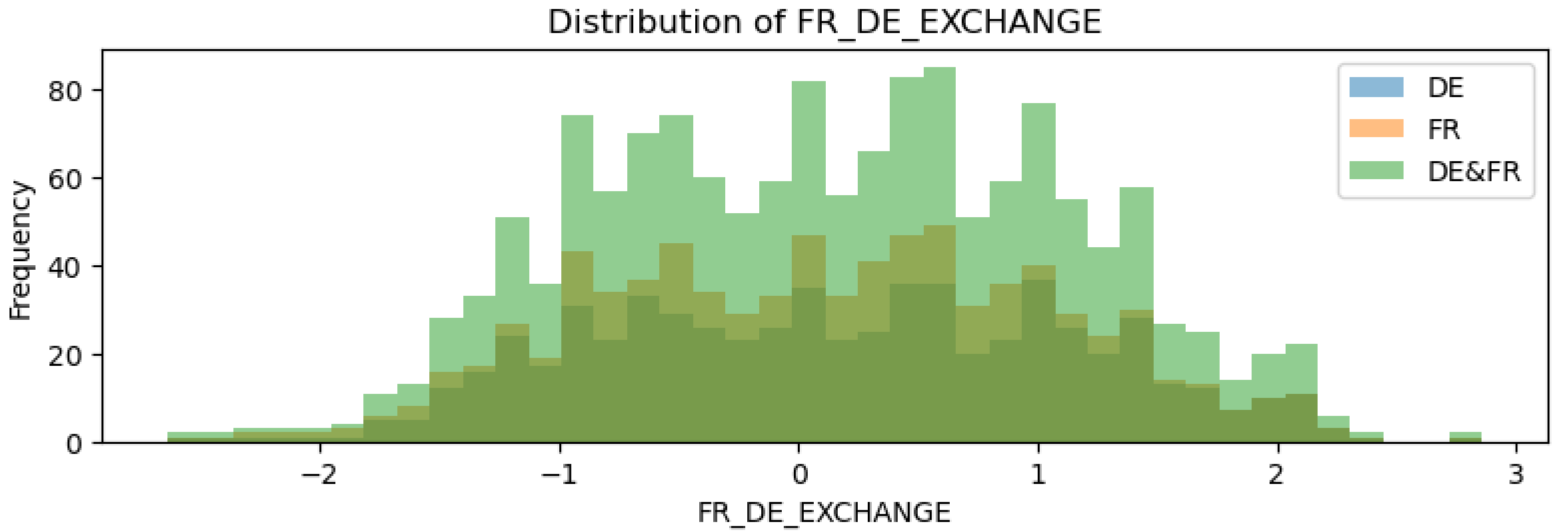
4. Distributions of Features

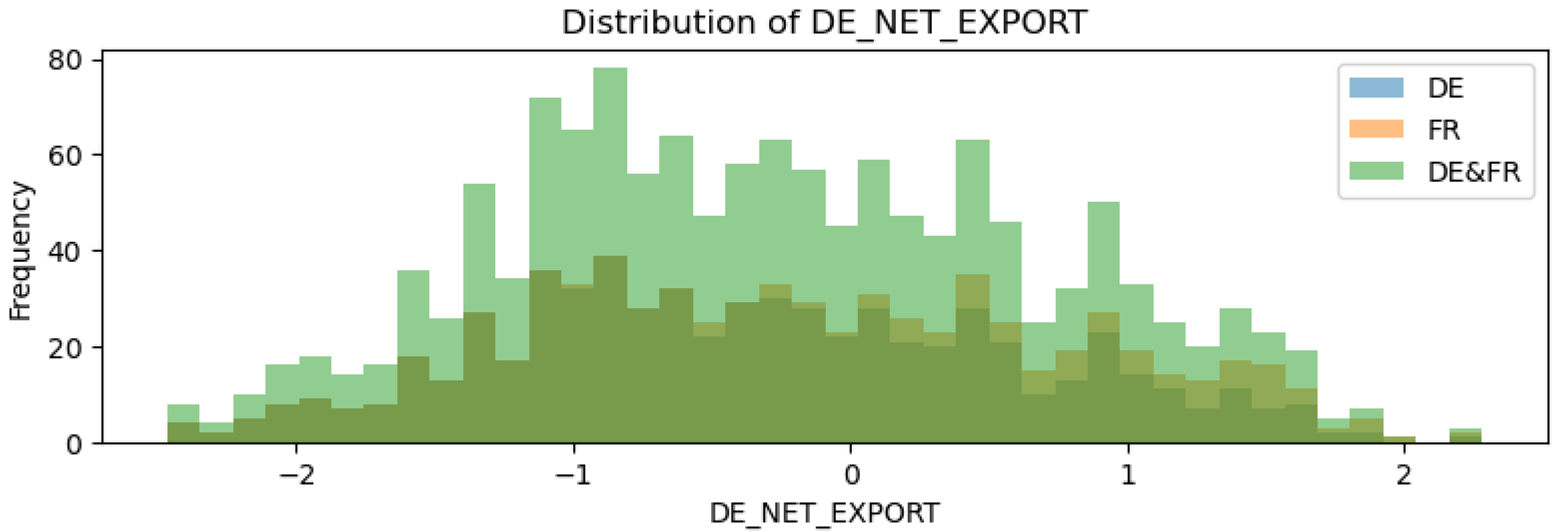
In [8]:
```
plot_cols = num_cols[1:]
for c in plot_cols:
    plt.figure()
    plt.subplot(2, 1, 1)
    if "COUNTRY" in df.columns:
        for k, g in df.groupby("COUNTRY"):
            g[c].plot(kind="hist", bins=40, alpha=0.5, label=str(k))
        df[c].plot(kind="hist", bins=40, alpha=0.5, label='DE&FR')
        plt.legend()
    else:
        df[c].plot(kind="hist", bins=40, alpha=0.7, shade=True)
    plt.xlabel(c)
    plt.title(f"Distribution of {c}")
    plt.tight_layout()
    plt.show()
```

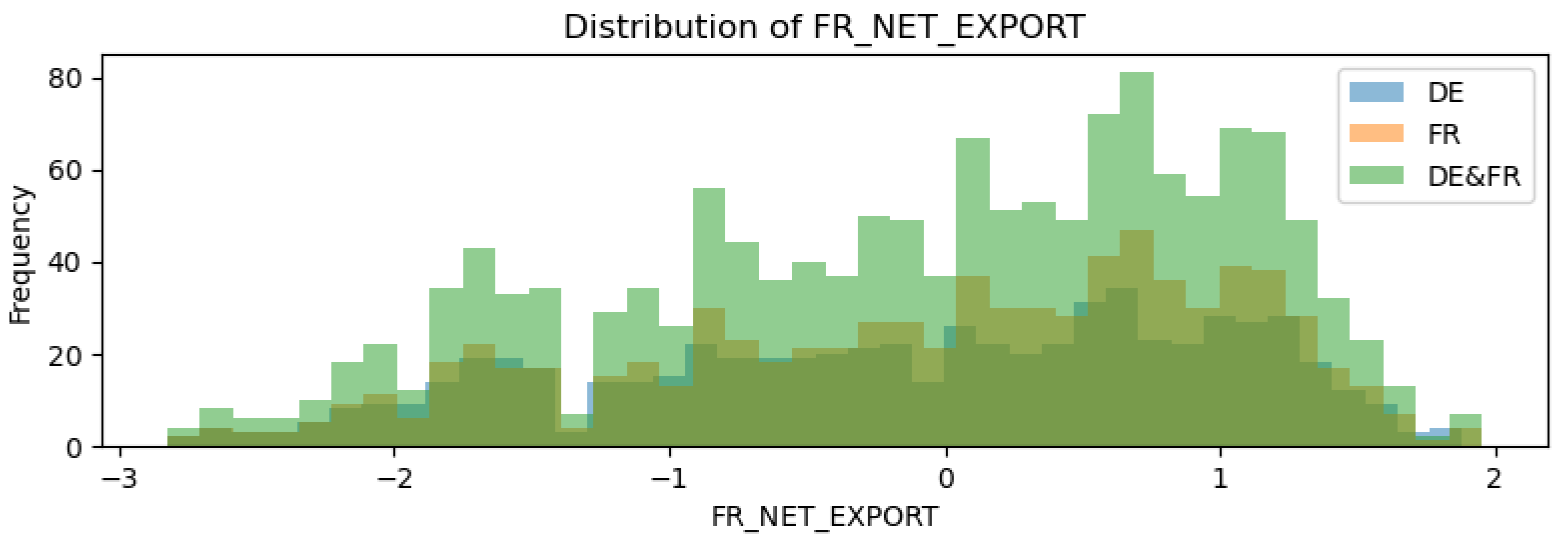


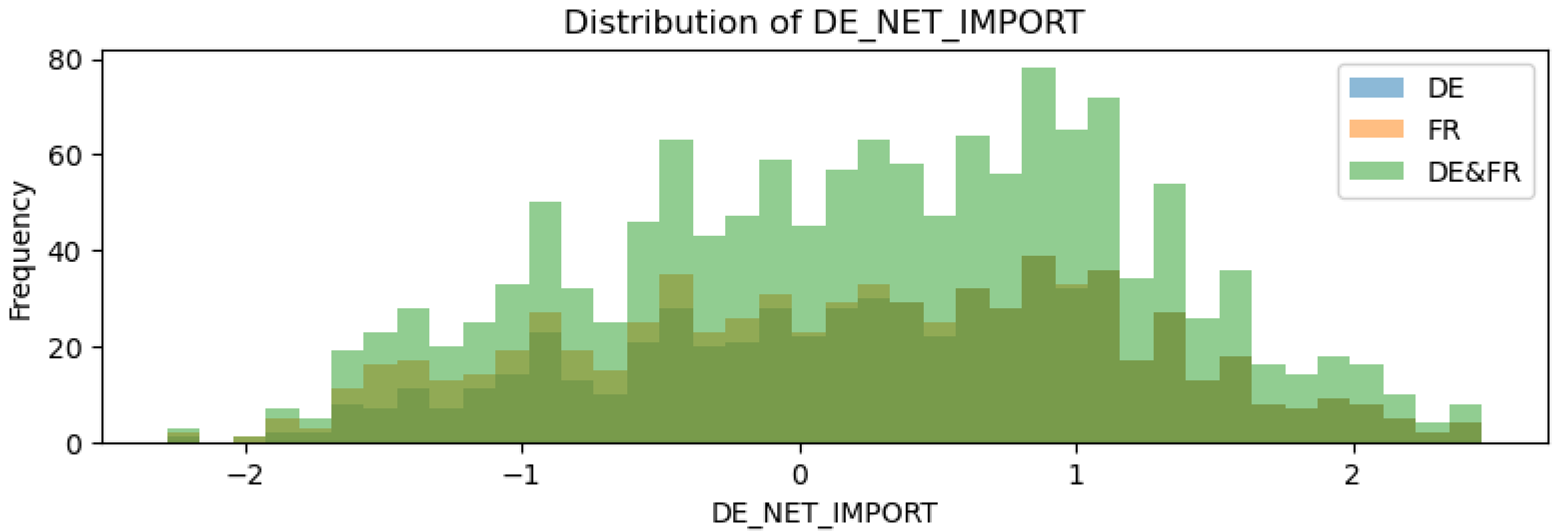

Distribution of FR_CONSUMPTION
Frequency
200
150
100
50
0
−1
0
1
2
3
FR_CONSUMPTION
DE
FR
DE&FR

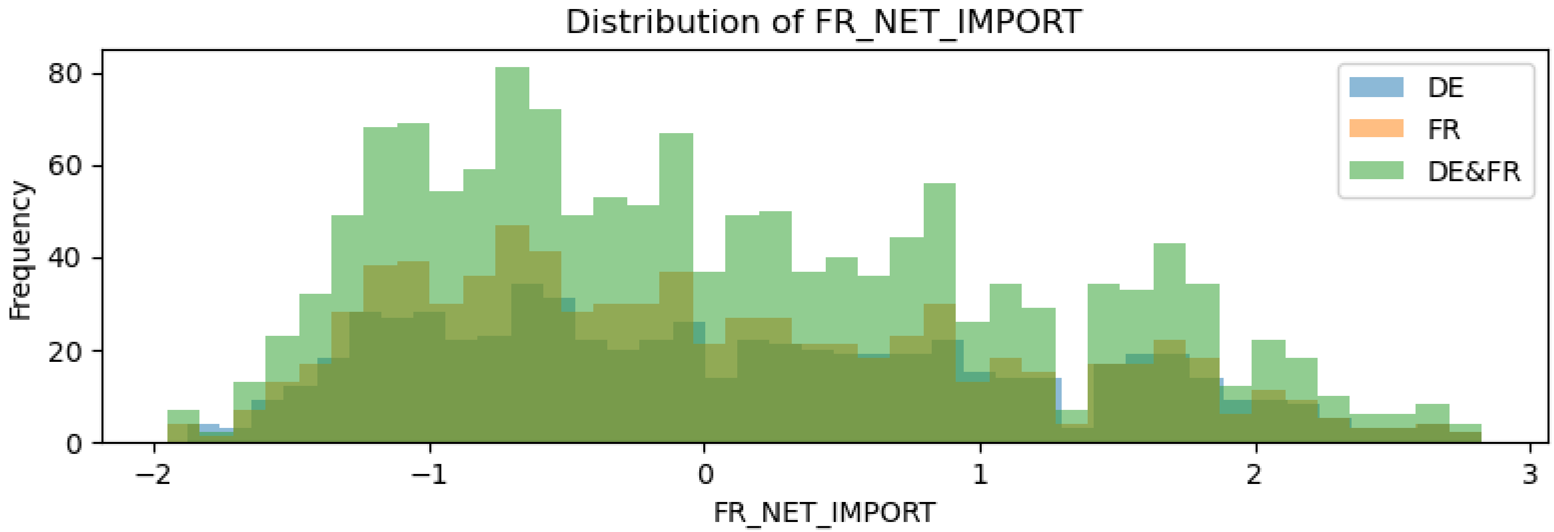

Distribution of DE_FR_EXCHANGE
Frequency
80
60
40
20
0
−3
−2
−1
0
1
2
DE_FR_EXCHANGE
DE
FR
DE&FR

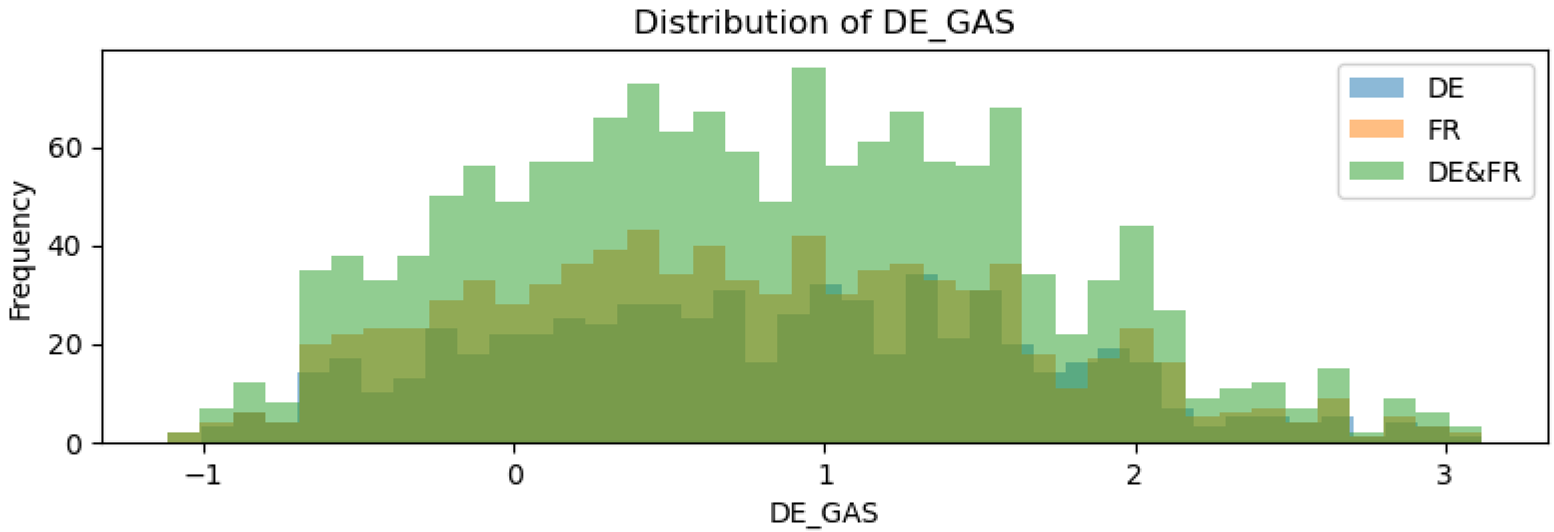

Distribution of FR_DE_EXCHANGE
Frequency
FR_DE_EXCHANGE
DE
FR
DE&FR
0
20
40
60
80
−2
−1
0
1
2
3

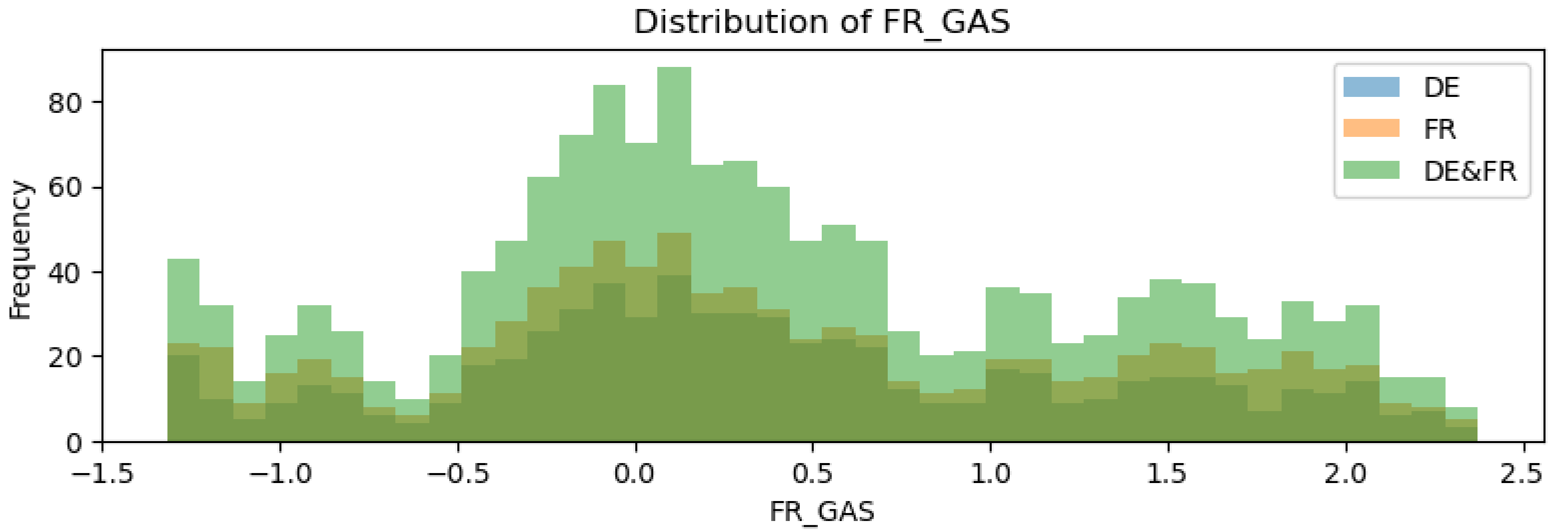

Distribution of DE_NET_EXPORT
Frequency
DE_NET_EXPORT
DE
FR
DE&FR
0
20
40
60
80
−2
−1
0
1
2

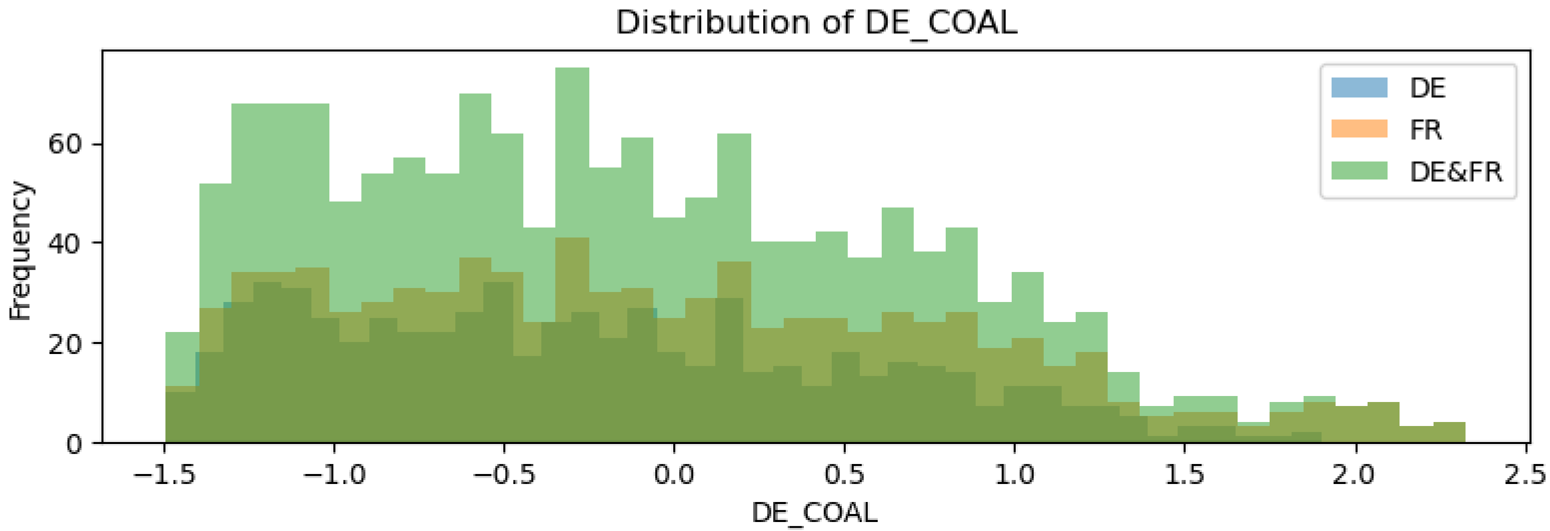

Distribution of FR_NET_EXPORT
Frequency
FR_NET_EXPORT
DE
FR
DE&FR
0
20
40
60
80
−3
−2
−1
0
1
2

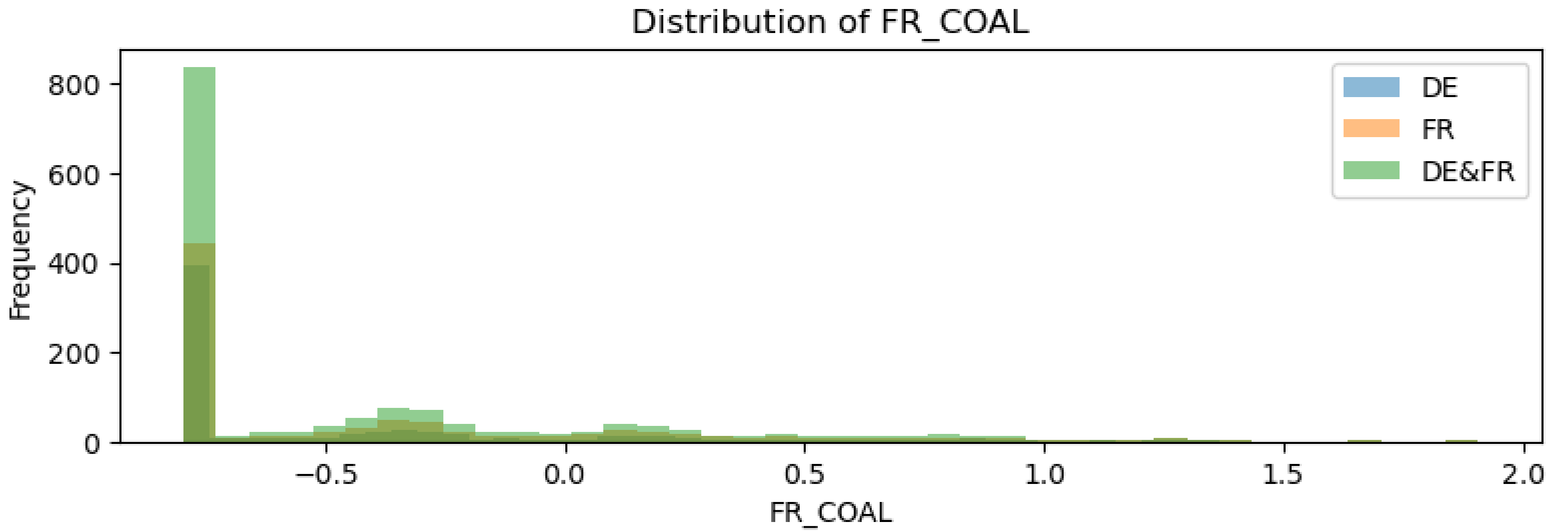

Distribution of DE_NET_IMPORT
Frequency
DE_NET_IMPORT
DE
FR
DE&FR
0
20
40
60
80
−2
−1
0
1
2

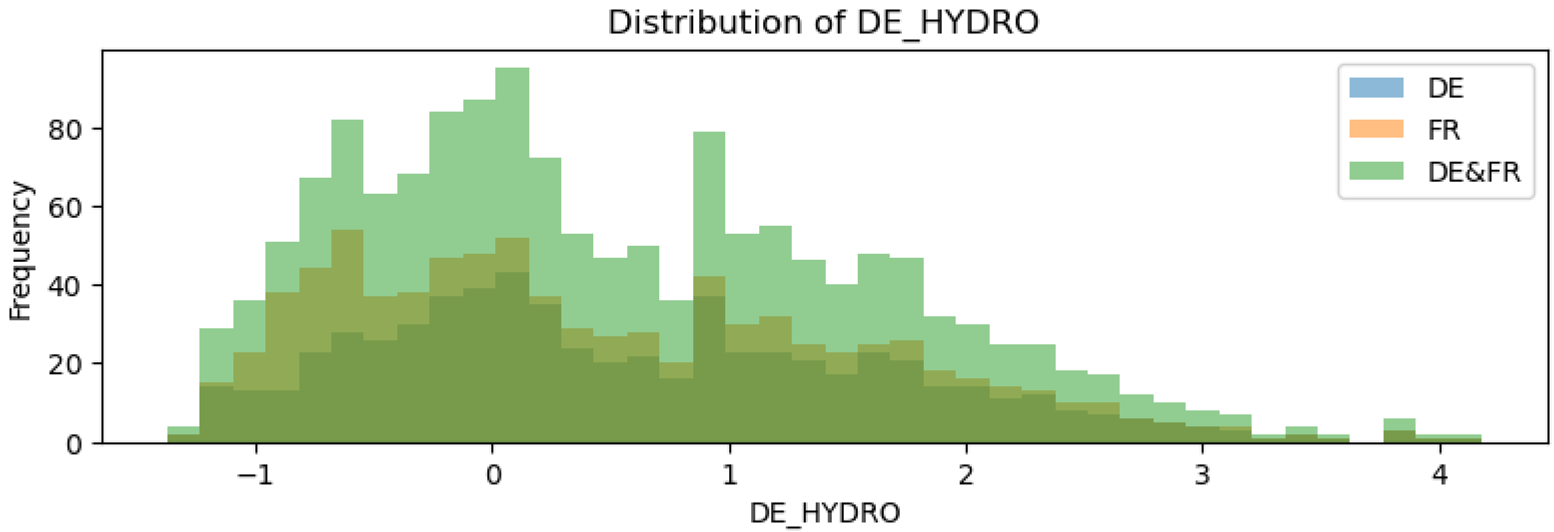

Distribution of FR_NET_IMPORT
80
60
40
20
0
Frequency
−2
−1
0
1
2
3
FR_NET_IMPORT
DE
FR
DE&FR

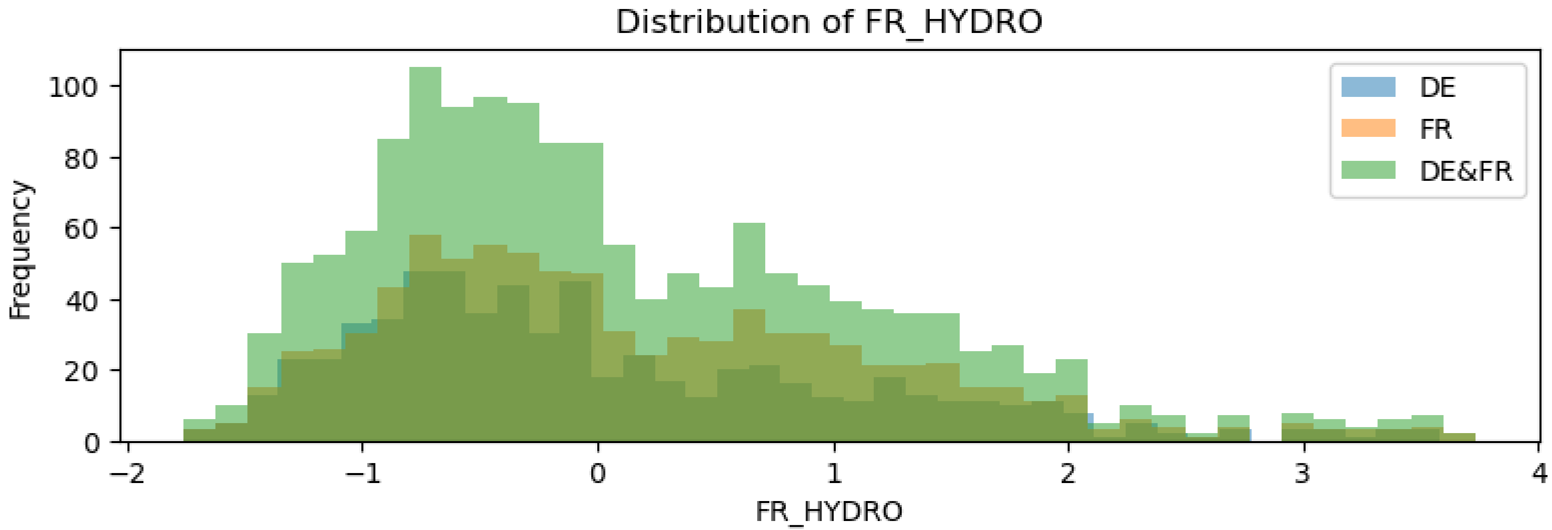

Distribution of DE_GAS
60
40
20
0
Frequency
−1
0
1
2
3
DE_GAS
DE
FR
DE&FR

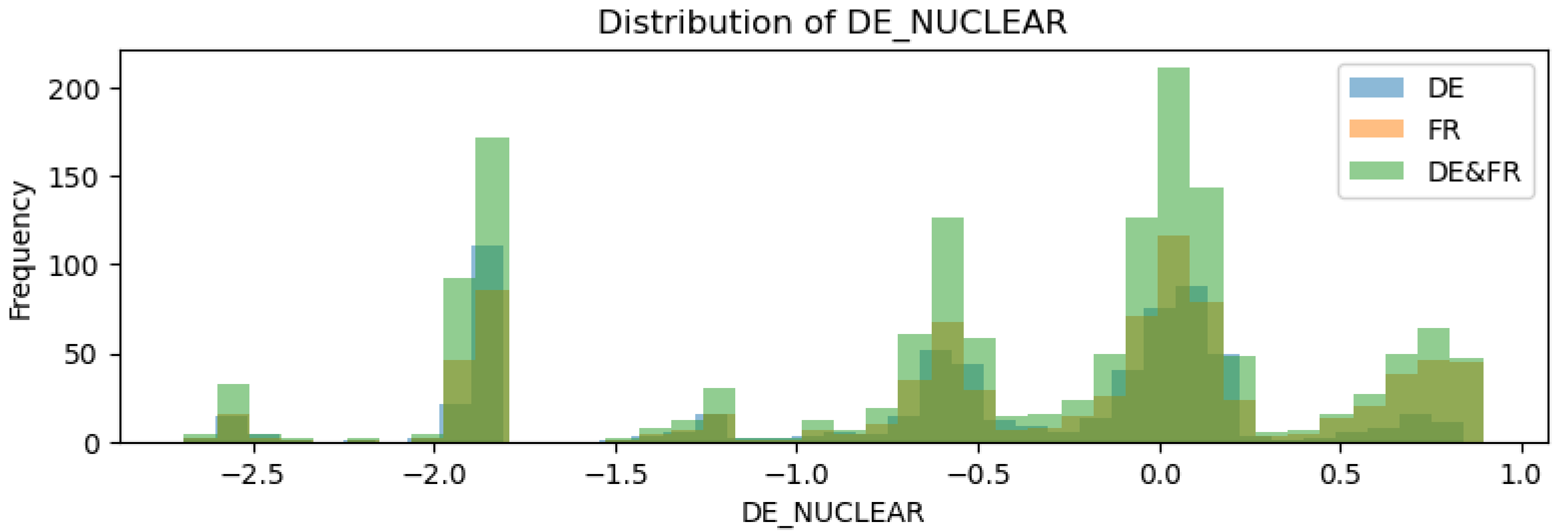

Distribution of FR_GAS
Frequency
FR_GAS
DE
FR
DE&FR
0
20
40
60
80
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5

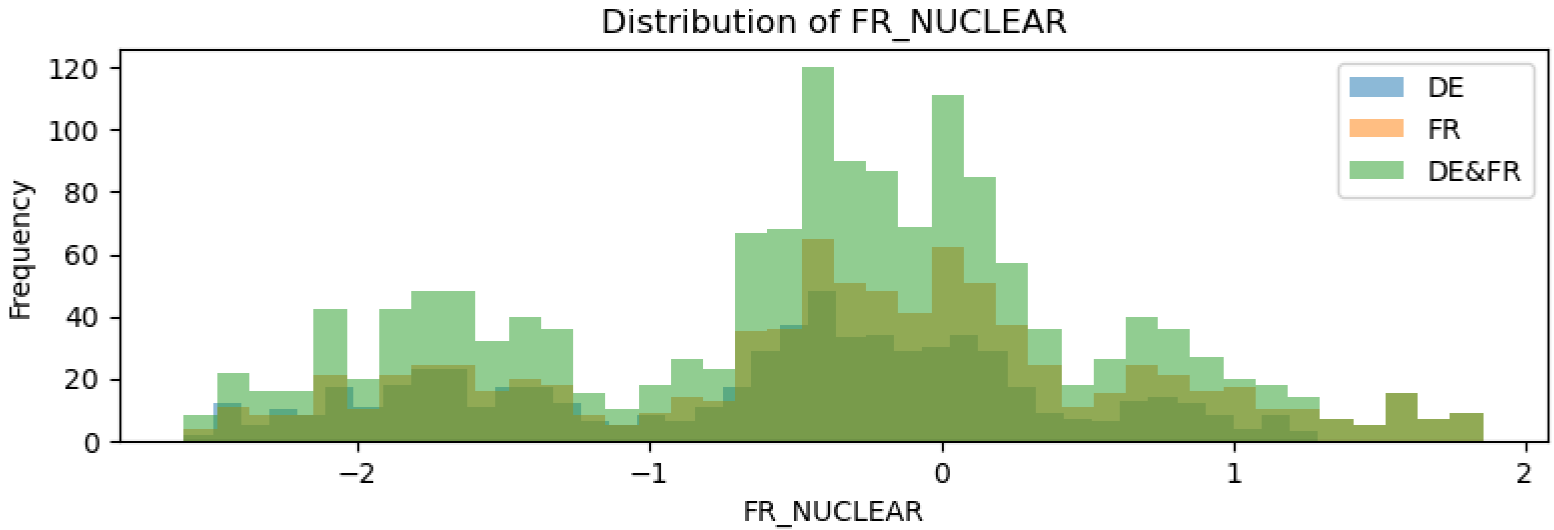

Distribution of DE_COAL
Frequency
DE_COAL
DE
FR
DE&FR
0
20
40
60
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5

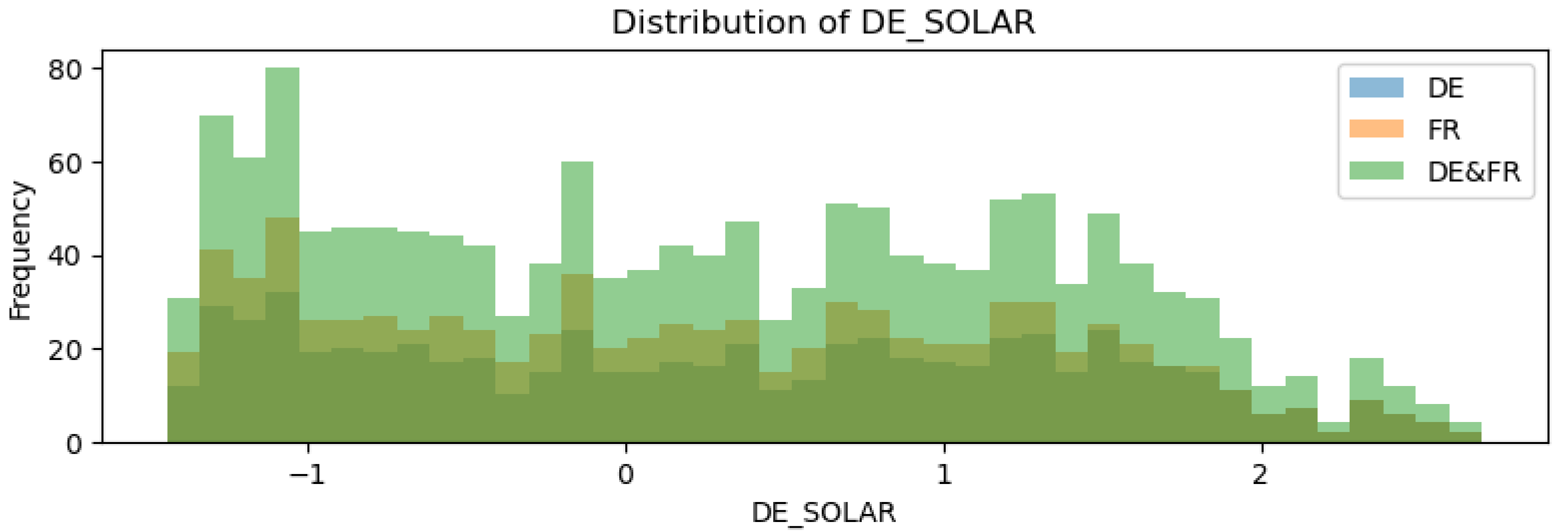

Distribution of FR_COAL
Frequency
800
600
400
200
0
−0.5
0.0
0.5
1.0
1.5
2.0
FR_COAL
DE
FR
DE&FR

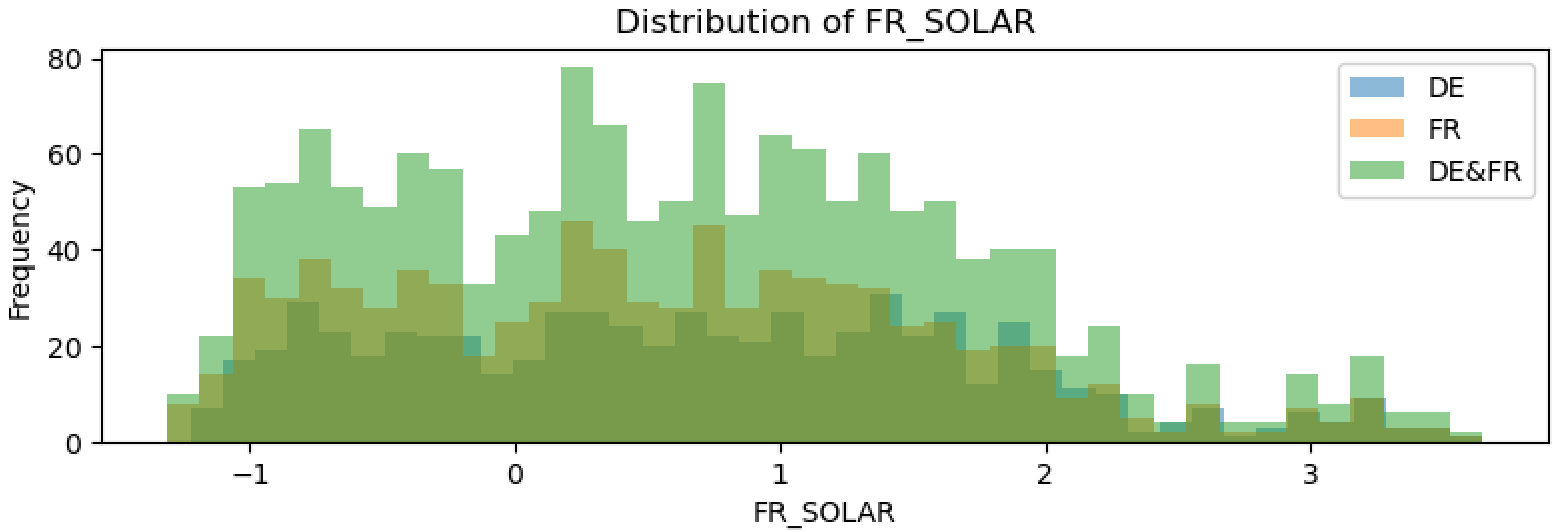

Distribution of DE_HYDRO
Frequency
80
60
40
20
0
−1
0
1
2
3
4
DE_HYDRO
DE
FR
DE&FR

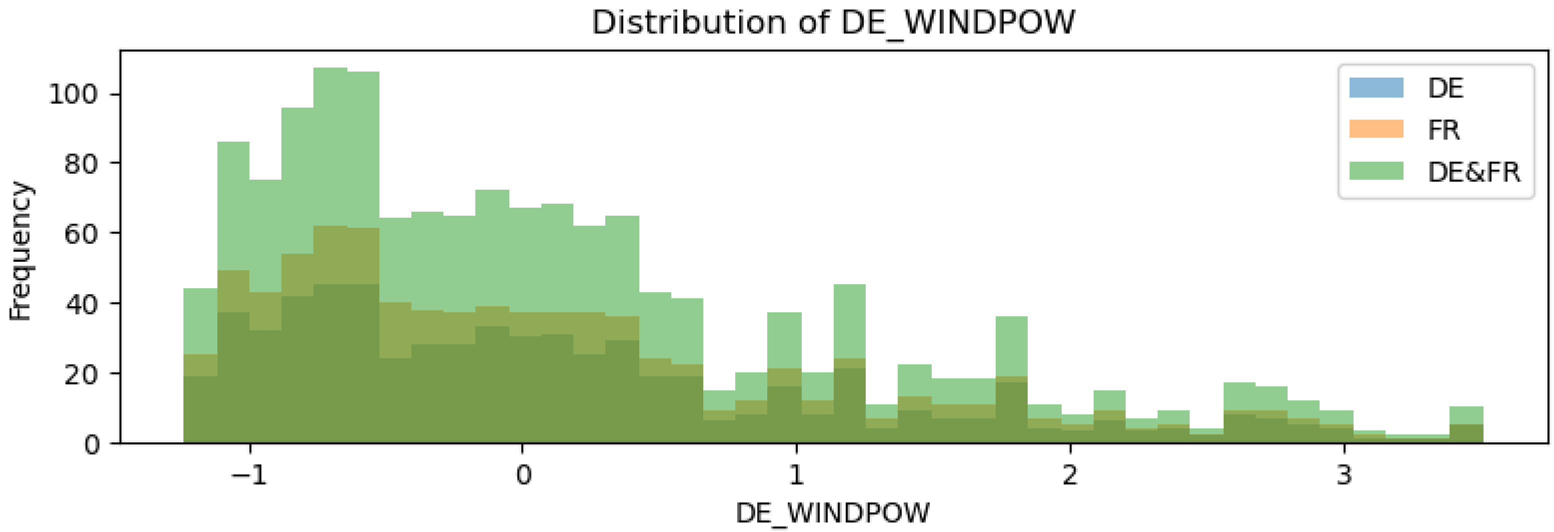

Distribution of FR_HYDRO
Frequency
FR_HYDRO
DE
FR
DE&FR
0
20
40
60
80
100
−2
−1
0
1
2
3
4

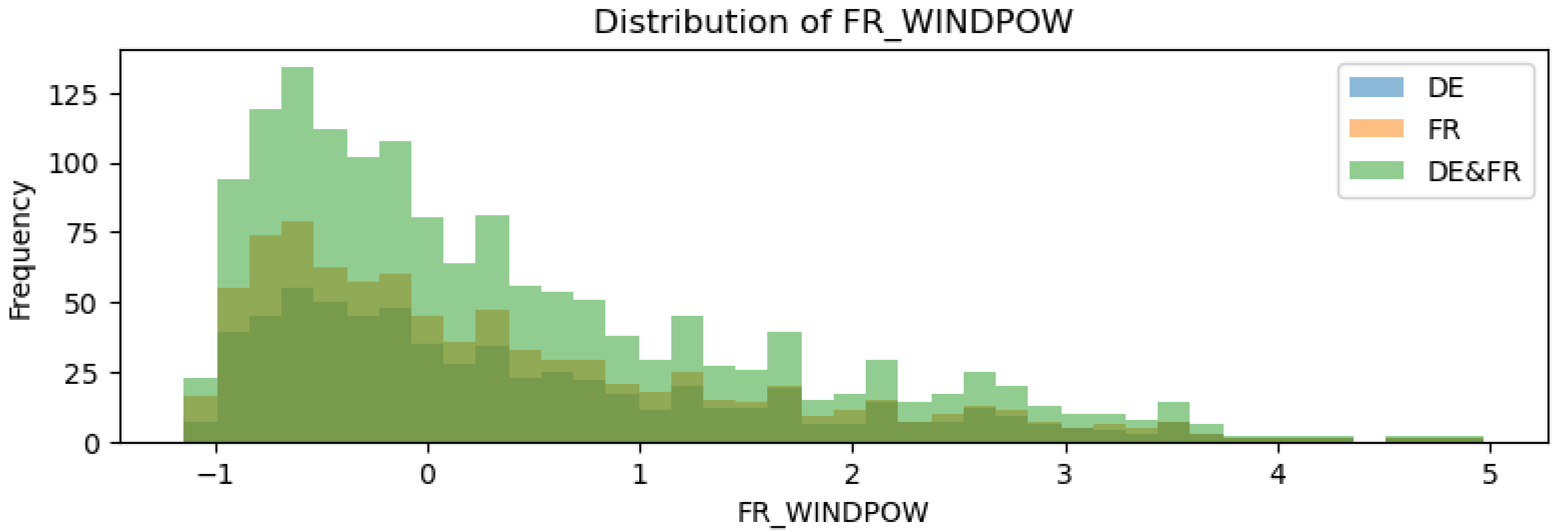

Distribution of DE_NUCLEAR
Frequency
DE_NUCLEAR
DE
FR
DE&FR
0
50
100
150
200
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0

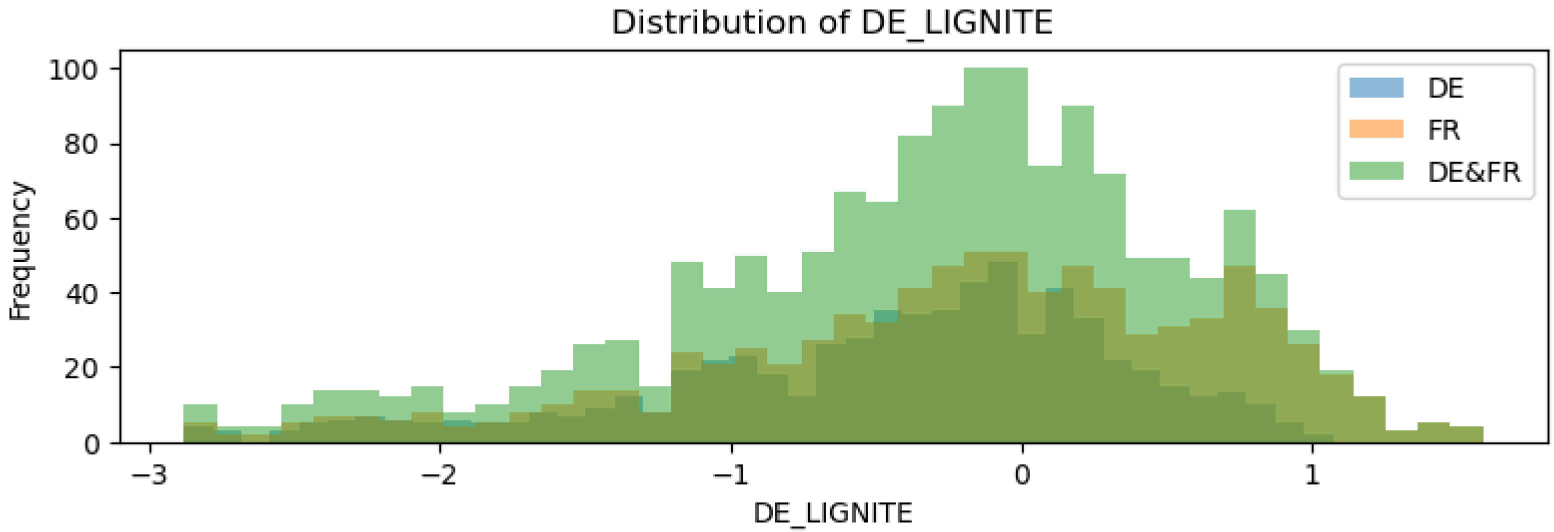

Distribution of FR_NUCLEAR
Frequency
FR_NUCLEAR
DE
FR
DE&FR

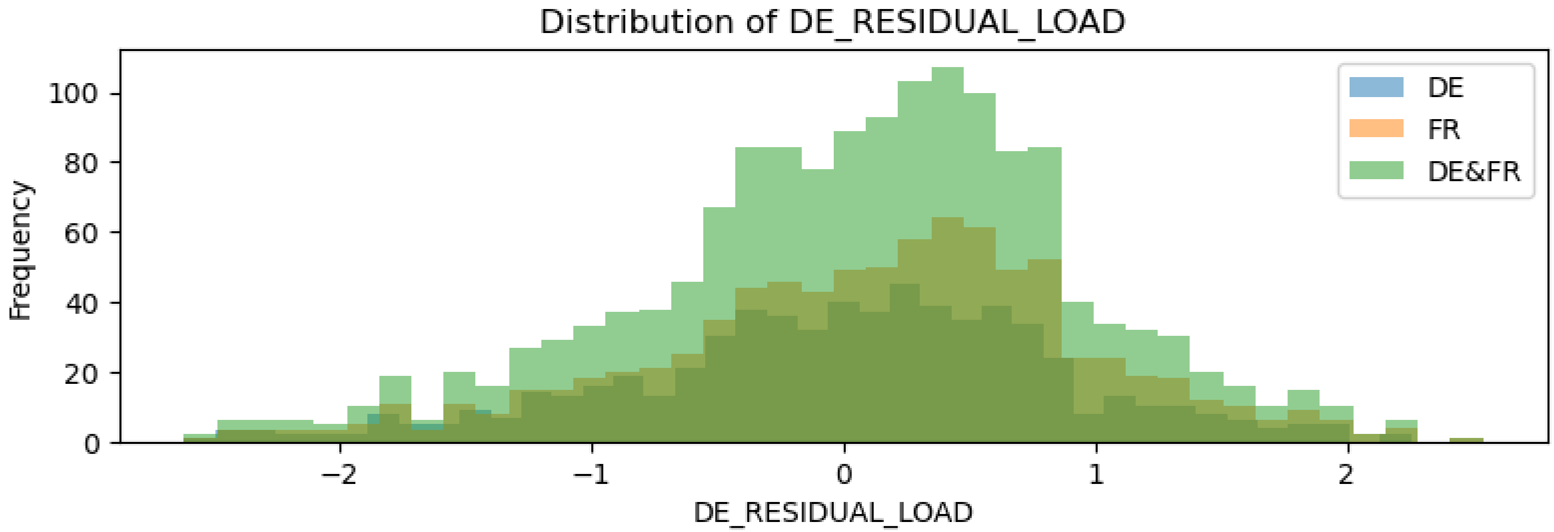

Distribution of DE_SOLAR
Frequency
DE_SOLAR
DE
FR
DE&FR

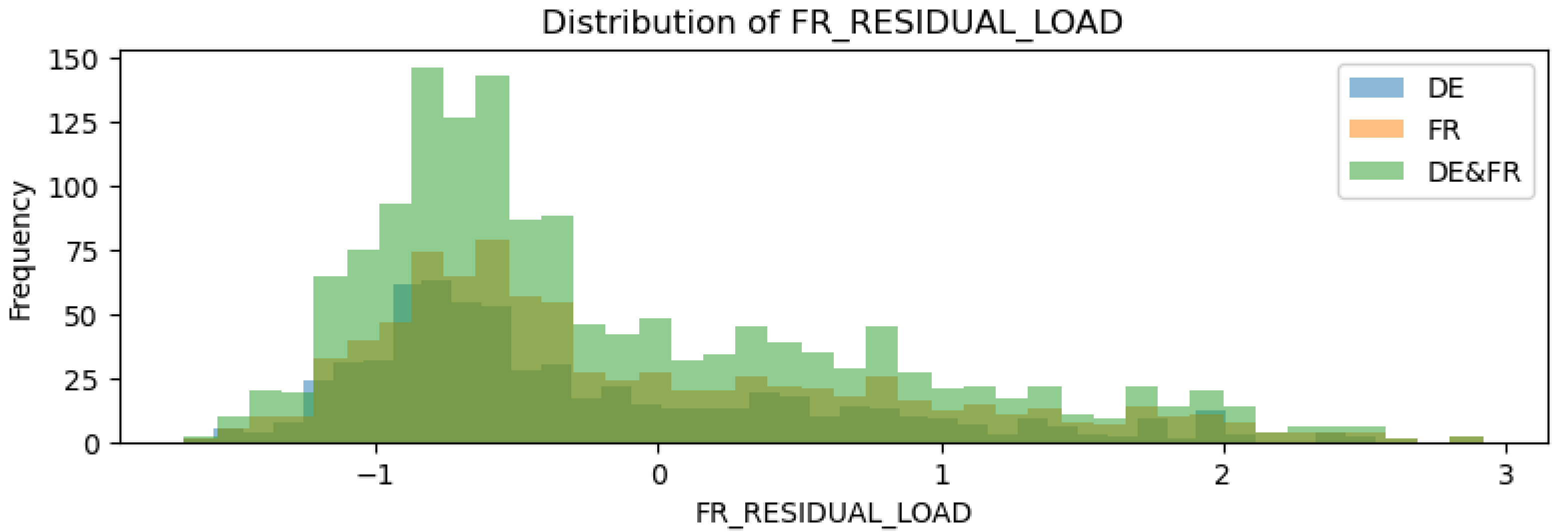

Distribution of FR_SOLAR
Frequency
0
20
40
60
80
−1
0
1
2
3
FR_SOLAR
DE
FR
DE&FR

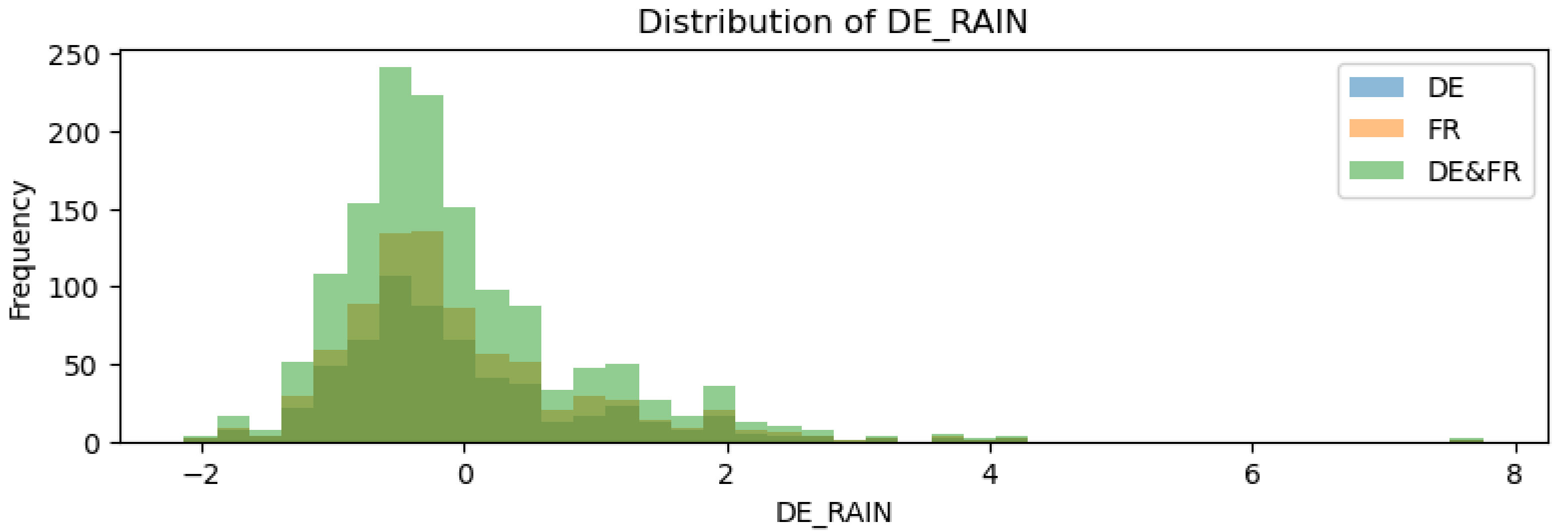

Distribution of DE_WINDPOW
Frequency
0
20
40
60
80
100
−1
0
1
2
3
DE_WINDPOW
DE
FR
DE&FR

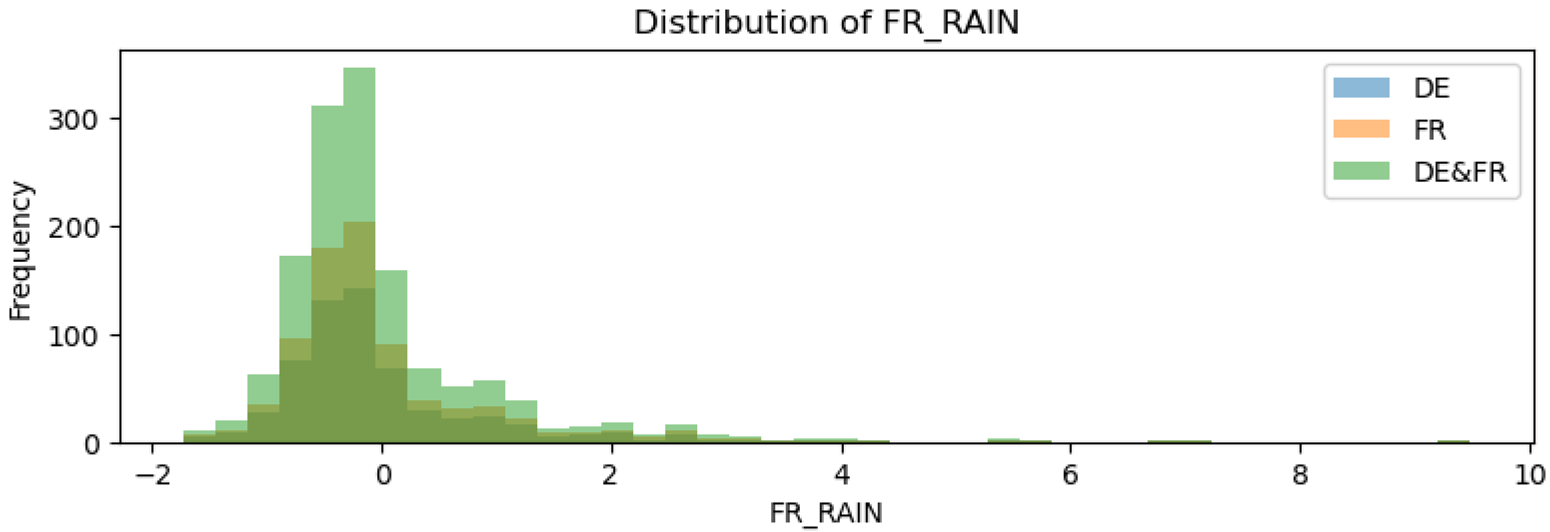

Distribution of FR_WINDPOW
Frequency
125
100
75
50
25
0
−1
0
1
2
3
4
5
FR_WINDPOW
DE
FR
DE&FR

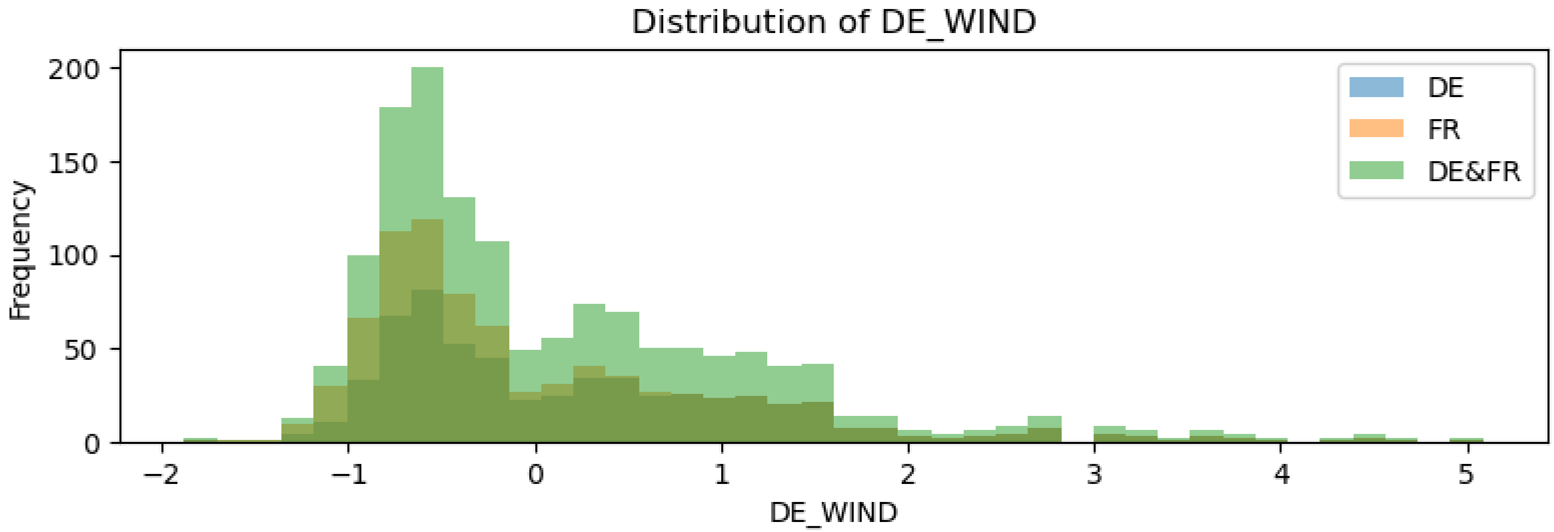

Distribution of DE_LIGNITE
Frequency
100
80
60
40
20
0
−3
−2
−1
0
1
DE_LIGNITE
DE
FR
DE&FR

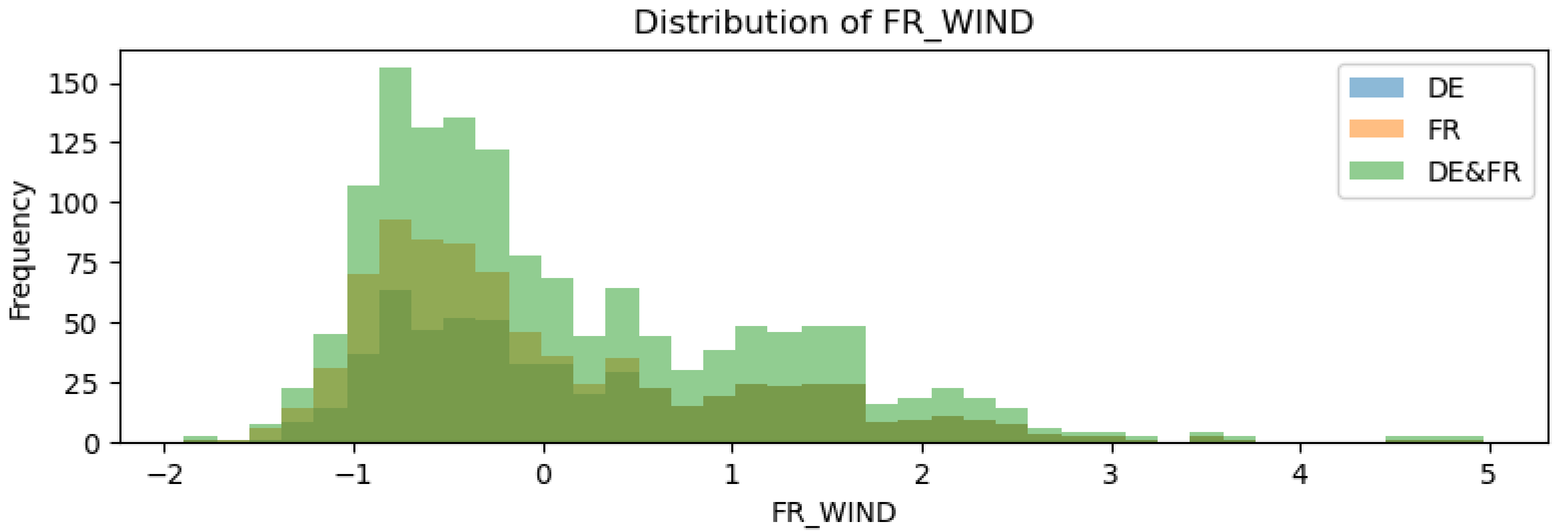

Distribution of DE_RESIDUAL_LOAD
Frequency
DE_RESIDUAL_LOAD
DE
FR
DE&FR
0
20
40
60
80
100
−2
−1
0
1
2

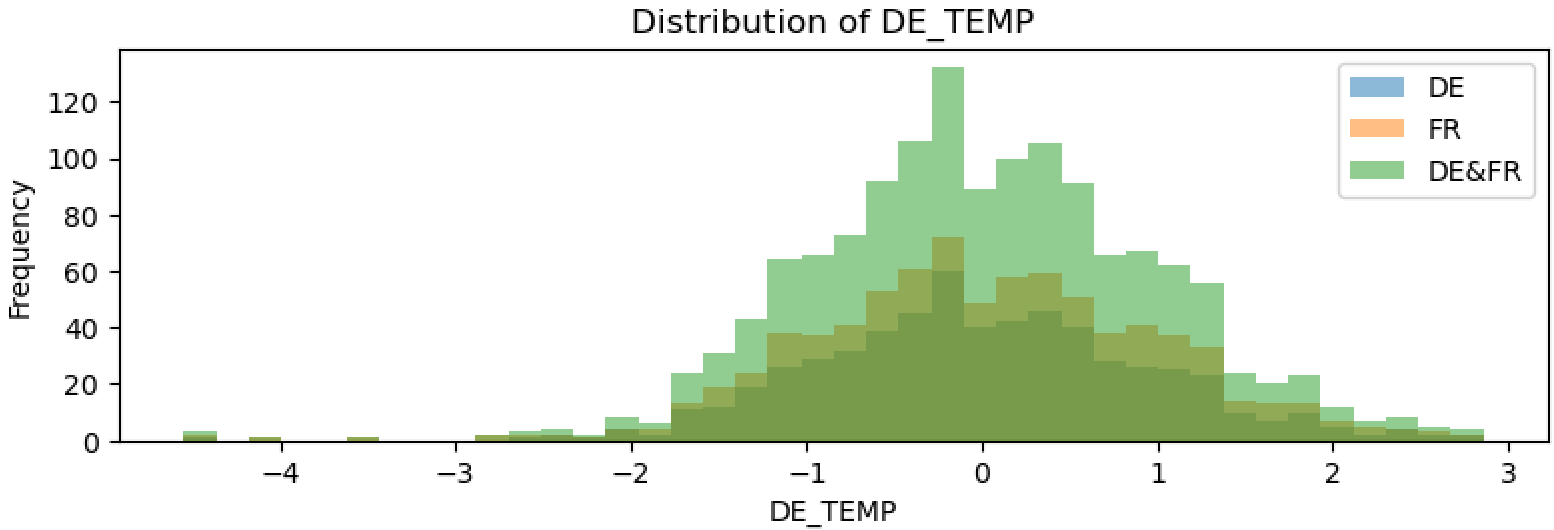

Distribution of FR_RESIDUAL_LOAD
Frequency
FR_RESIDUAL_LOAD
DE
FR
DE&FR
0
25
50
75
100
125
150
−1
0
1
2
3

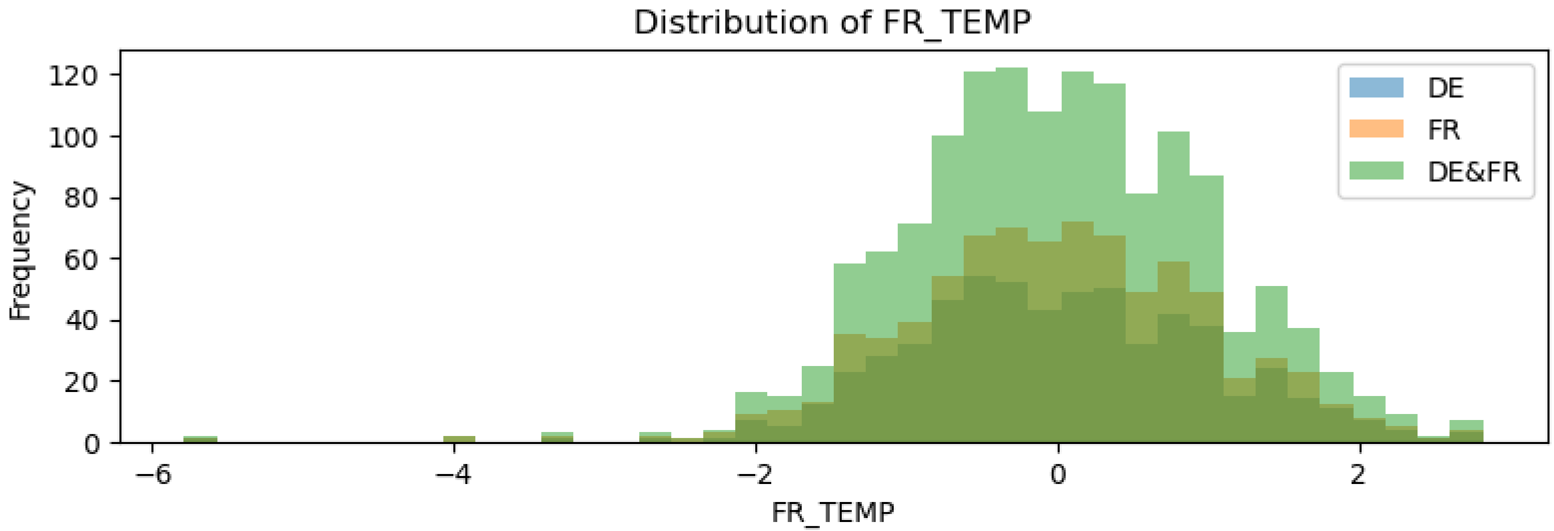

Distribution of DE_RAIN
Frequency
0
50
100
150
200
250
−2
0
2
4
6
8
DE_RAIN
DE
FR
DE&FR

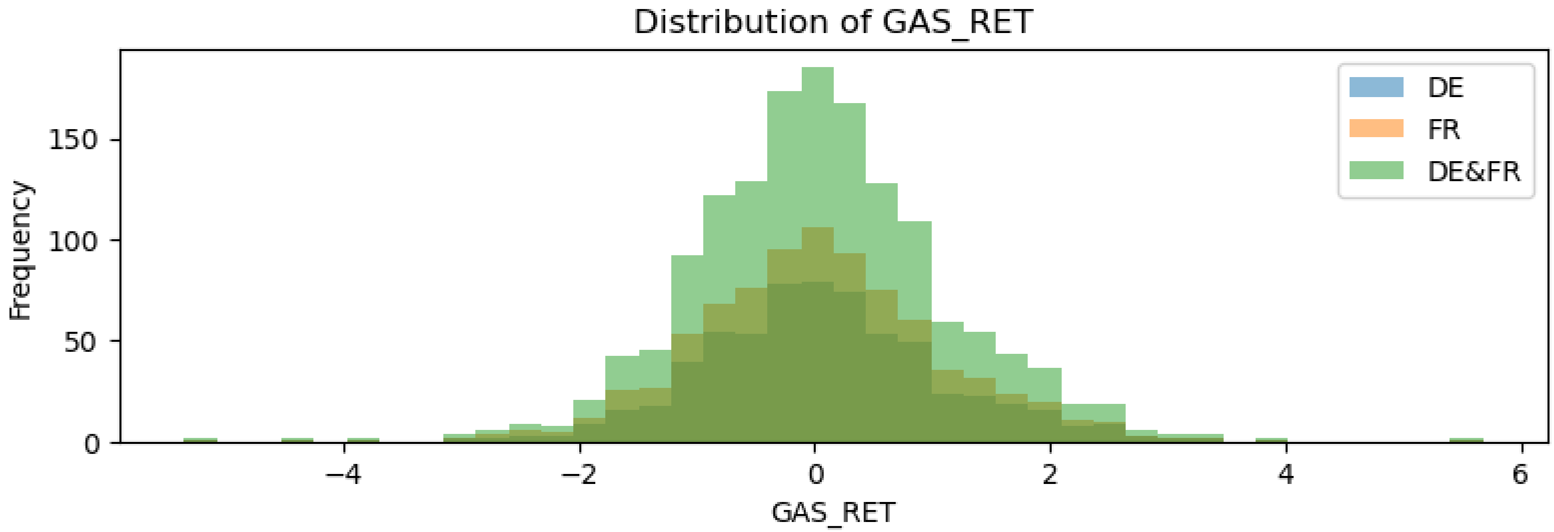

Distribution of FR_RAIN
Frequency
0
100
200
300
−2
0
2
4
6
8
10
FR_RAIN
DE
FR
DE&FR

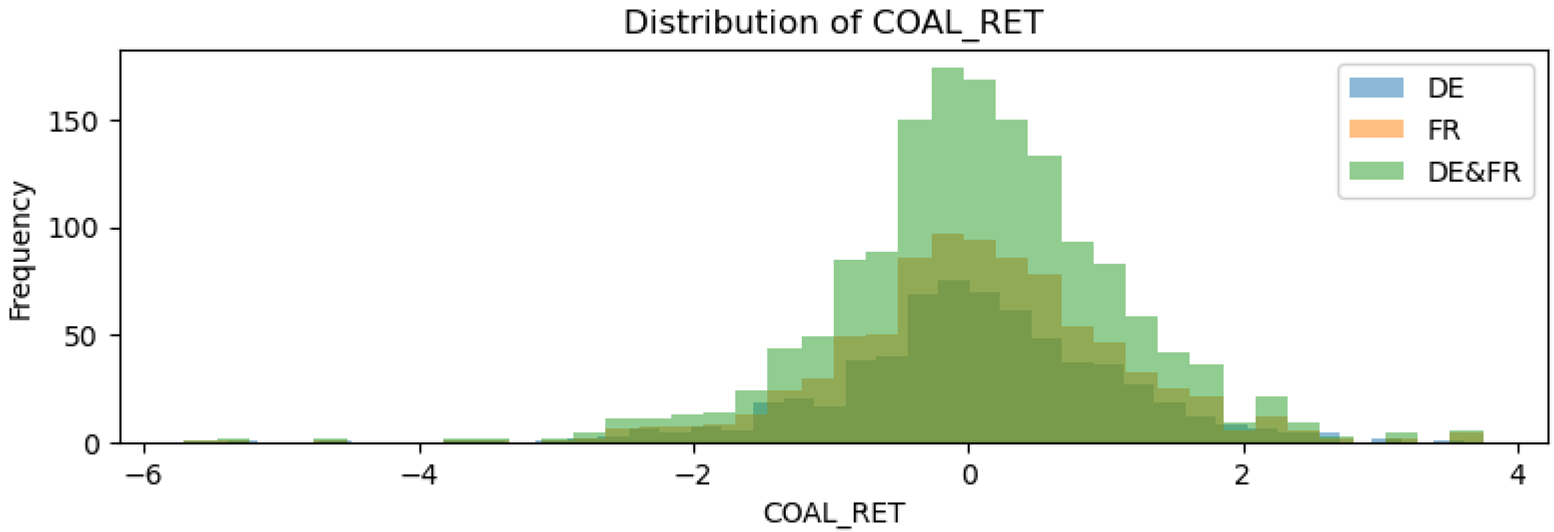

Distribution of DE_WIND
Frequency
DE_WIND
DE
FR
DE&FR
200
150
100
50
0
−2
−1
0
1
2
3
4
5

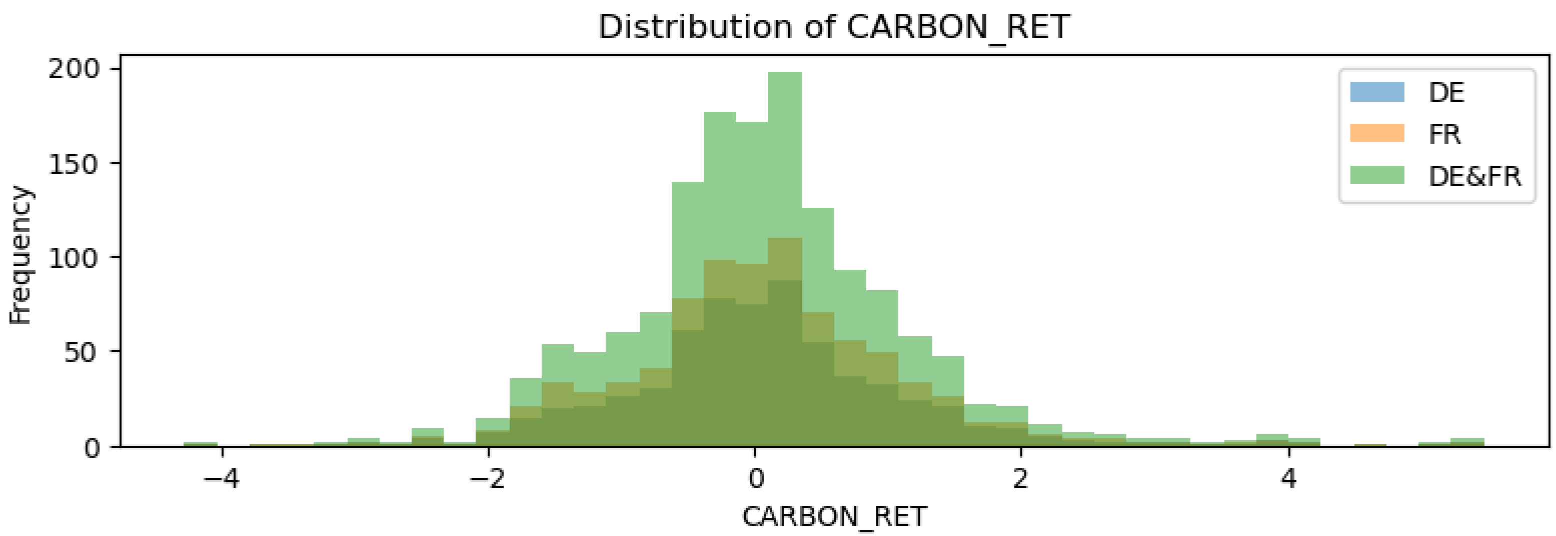

Distribution of FR_WIND
Frequency
FR_WIND
DE
FR
DE&FR
150
125
100
75
50
25
0
−2
−1
0
1
2
3
4
5


Distribution of DE_TEMP
Frequency
DE_TEMP
DE
FR
DE&FR
0
20
40
60
80
100
120
−4
−3
−2
−1
0
1
2
3


Distribution of FR_TEMP
Frequency
FR_TEMP
DE
FR
DE&FR
0
20
40
60
80
100
120
−6
−4
−2
0
2


Distribution of GAS_RET
Frequency
150
100
50
0
−4
−2
0
2
4
6
GAS_RET
DE
FR
DE&FR

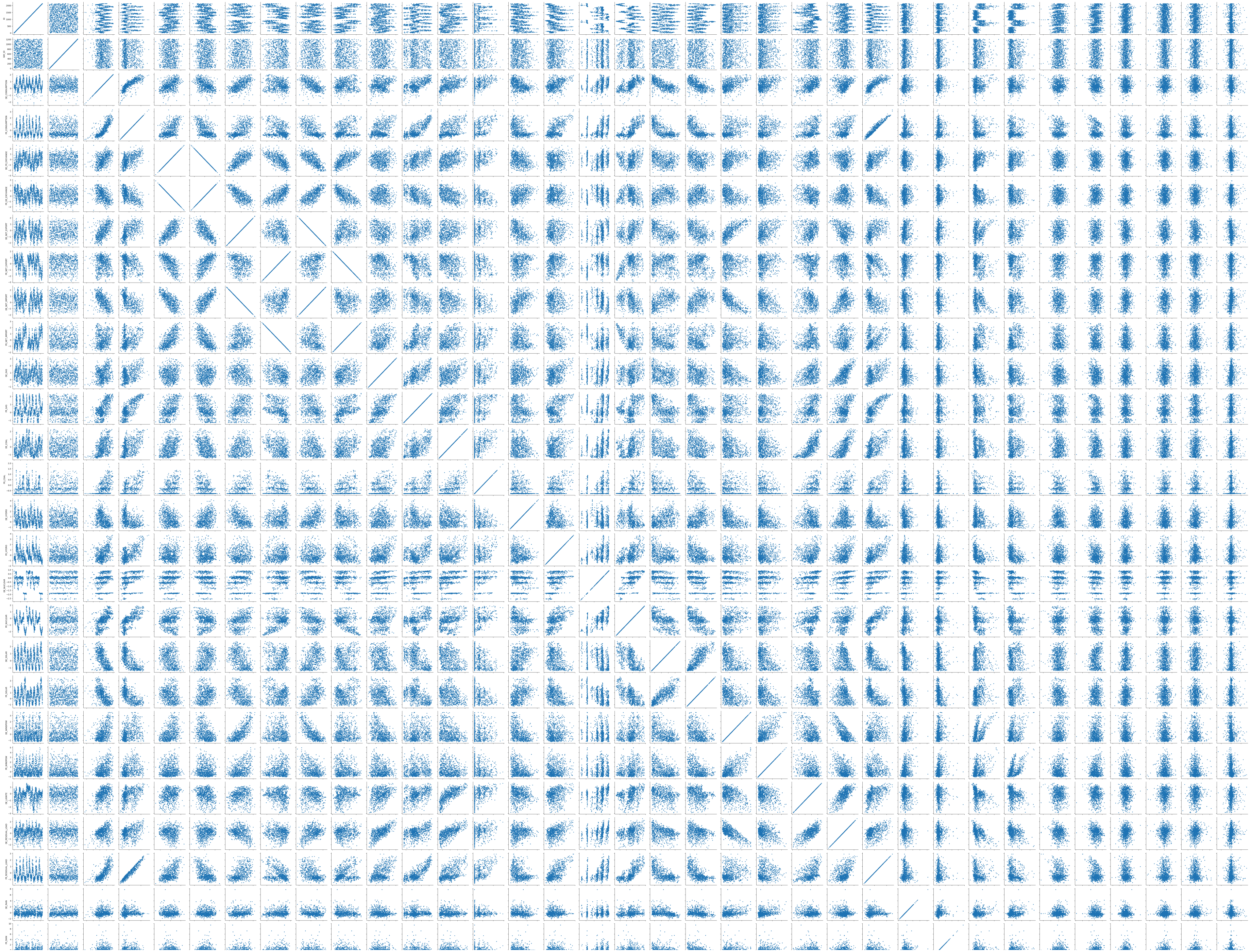

Distribution of COAL_RET
Frequency
150
100
50
0
−6
−4
−2
0
2
4
COAL_RET
DE
FR
DE&FR

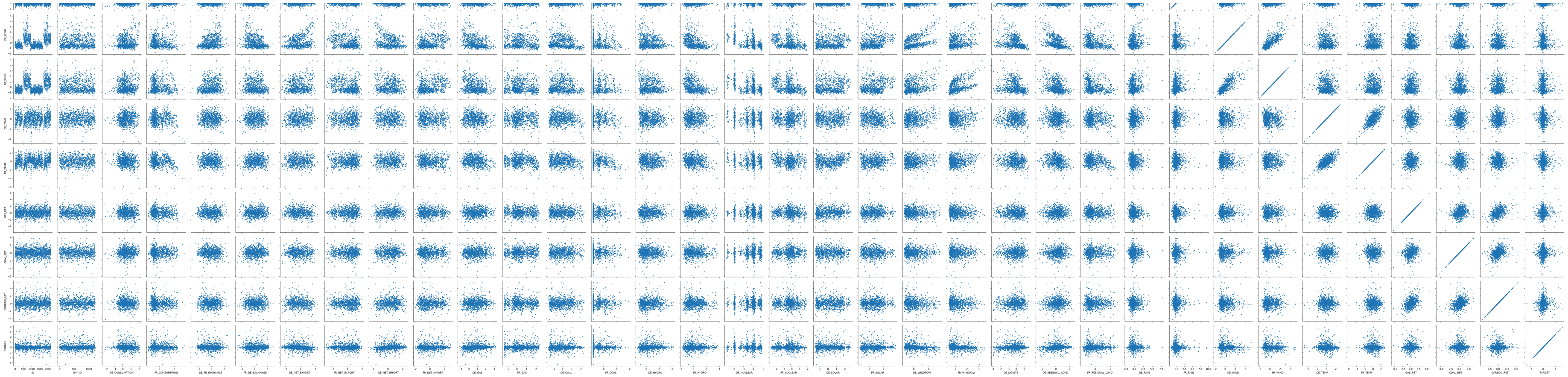


5. Feature-Feature Correlations

In [9]:
```
sns.PairGrid(X.merge(y, on='ID', how='left')).map(plt.scatter,s=5)
```

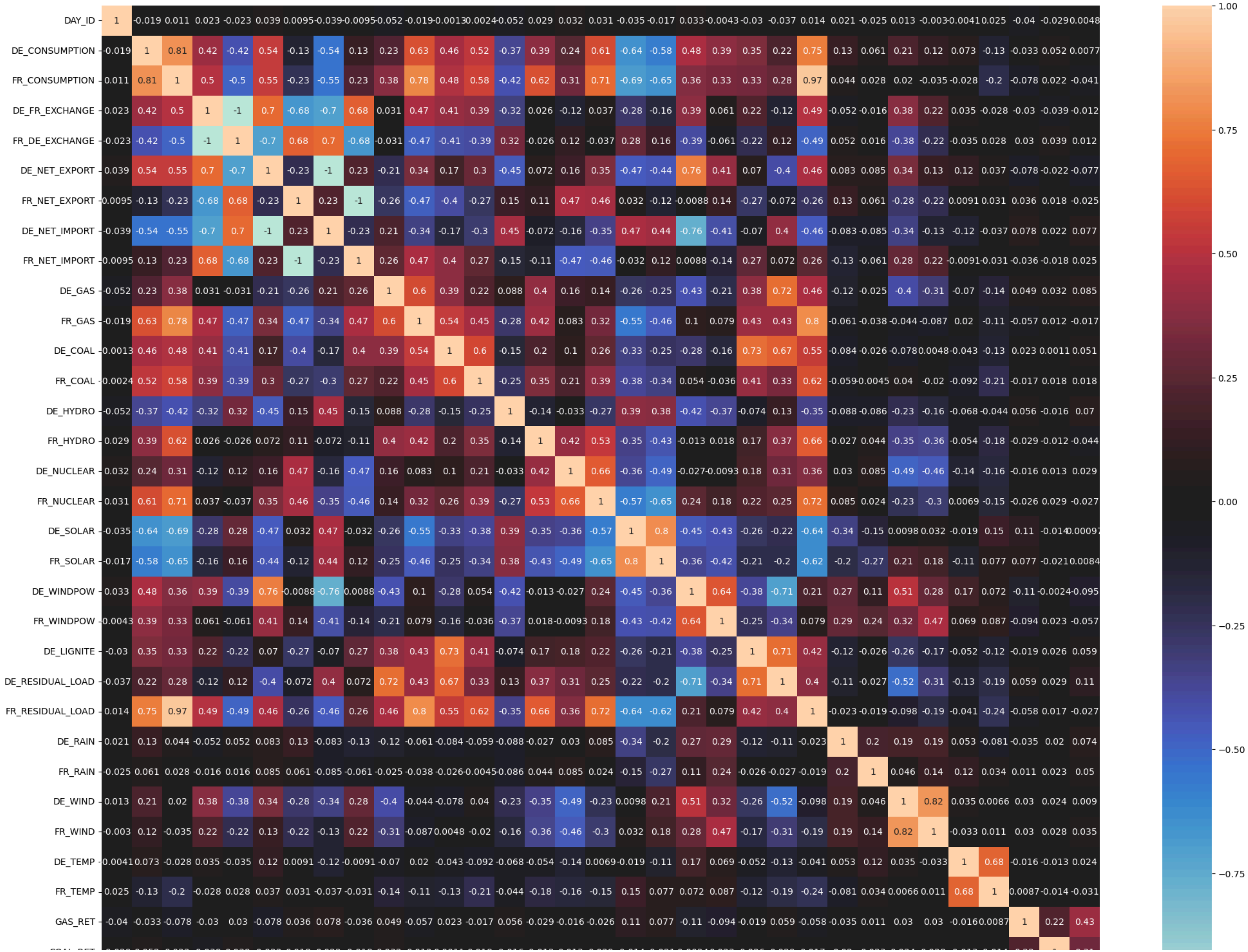
Out[9]: <seaborn.axisgrid.PairGrid at 0x2534b2da560>

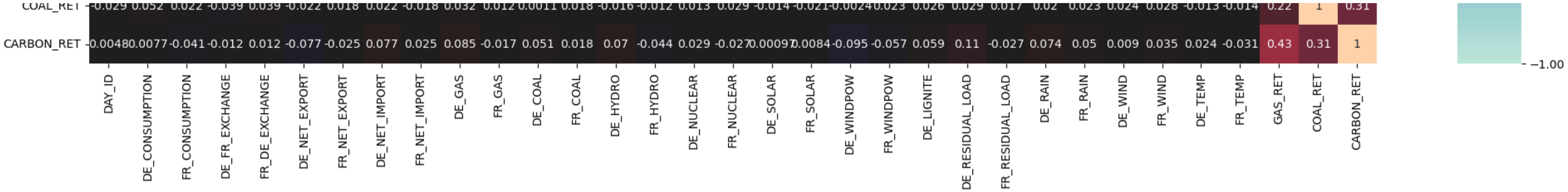




In [10]:
```
corr = df[num_cols].corr()
print("Correlation matrix shape:", corr.shape)

plt.figure(figsize=[25,20])
sns.heatmap(corr, annot=True, vmin=-1, vmax=1, center=0)
plt.show()
```

```
Correlation matrix shape: (33, 33)
```


DAY_ID
DE_CONSUMPTION
FR_CONSUMPTION
DE_FR_EXCHANGE
FR_DE_EXCHANGE
DE_NET_EXPORT
FR_NET_EXPORT
DE_NET_IMPORT
FR_NET_IMPORT
DE_GAS
FR_GAS
DE_COAL
FR_COAL
DE_HYDRO
FR_HYDRO
DE_NUCLEAR
FR_NUCLEAR
DE_SOLAR
FR_SOLAR
DE_WINDPOW
FR_WINDPOW
DE_LIGNITE
DE_RESIDUAL_LOAD
FR_RESIDUAL_LOAD
DE_RAIN
FR_RAIN
DE_WIND
FR_WIND
DE_TEMP
FR_TEMP
GAS_RET
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75



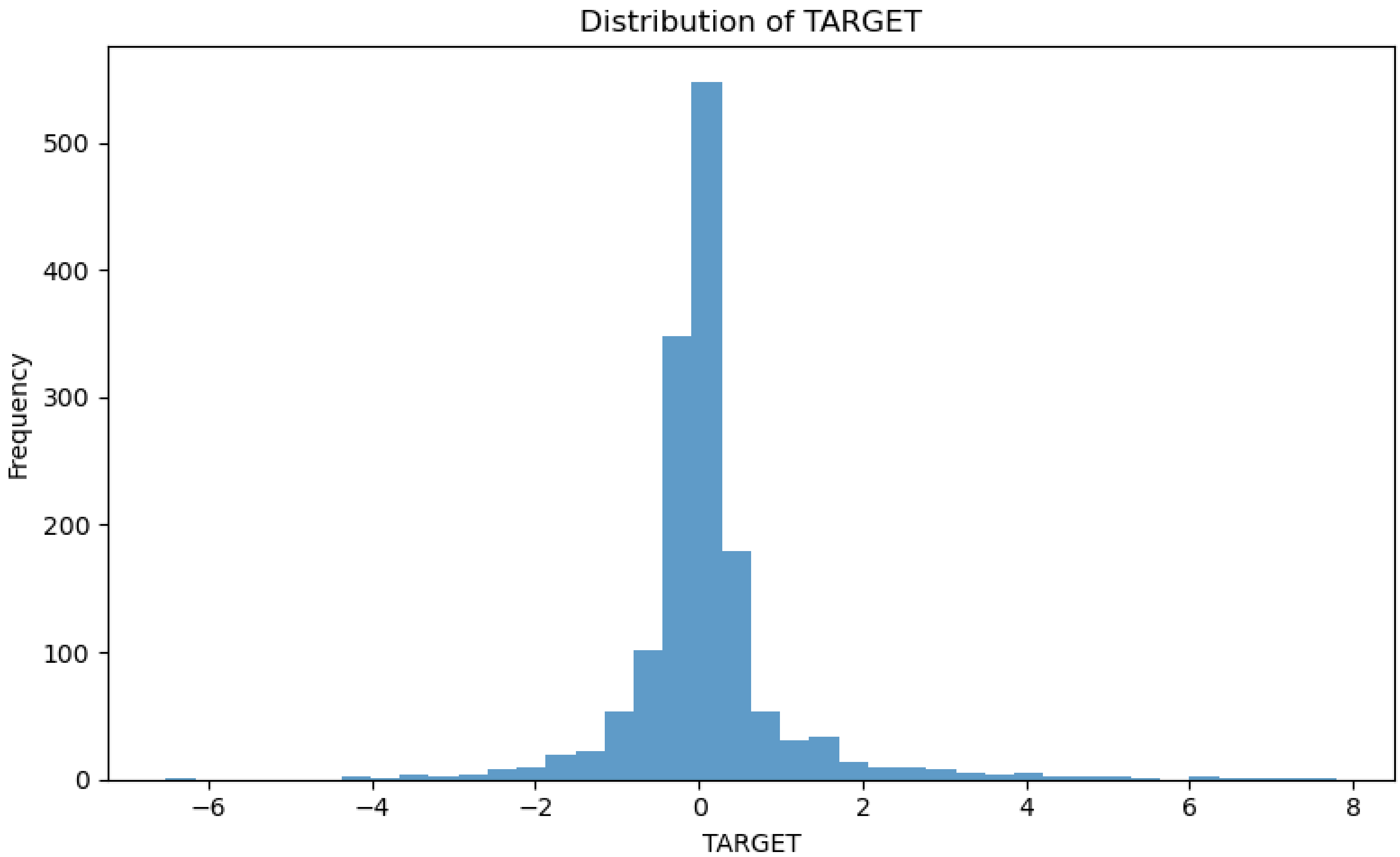
In [11]:
```python
# Top correlated pairs by |r|
pairs = []
cols = corr.columns
for i in range(len(cols)):
    for j in range(i+1, len(cols)):
        r = corr.iloc[i, j]
        pairs.append((cols[i], cols[j], abs(r), r))
pairs_sorted = sorted(pairs, key=lambda x: x[2], reverse=True)[:20]
pd.DataFrame(pairs_sorted, columns=["feat1", "feat2", "|r|", "r"])
```

Out[11]:

	feat1	feat2	|r|	r
0	DE_FR_EXCHANGE	FR_DE_EXCHANGE	1.000000	-1.000000
1	DE_NET_EXPORT	DE_NET_IMPORT	1.000000	-1.000000
2	FR_NET_EXPORT	FR_NET_IMPORT	1.000000	-1.000000
3	FR_CONSUMPTION	FR_RESIDUAL_LOAD	0.965086	0.965086
4	DE_WIND	FR_WIND	0.820724	0.820724
5	DE_CONSUMPTION	FR_CONSUMPTION	0.813202	0.813202
6	DE_SOLAR	FR_SOLAR	0.803381	0.803381
7	FR_GAS	FR_RESIDUAL_LOAD	0.795914	0.795914
8	FR_CONSUMPTION	FR_GAS	0.779727	0.779727
9	DE_NET_EXPORT	DE_WINDPOW	0.763398	0.763398
10	DE_NET_IMPORT	DE_WINDPOW	0.763398	-0.763398
11	DE_CONSUMPTION	FR_RESIDUAL_LOAD	0.754142	0.754142
12	DE_COAL	DE_LIGNITE	0.725934	0.725934
13	FR_NUCLEAR	FR_RESIDUAL_LOAD	0.717724	0.717724
14	DE_GAS	DE_RESIDUAL_LOAD	0.715224	0.715224
15	DE_LIGNITE	DE_RESIDUAL_LOAD	0.713916	0.713916
16	DE_WINDPOW	DE_RESIDUAL_LOAD	0.705836	-0.705836
17	FR_CONSUMPTION	FR_NUCLEAR	0.705734	0.705734
18	DE_FR_EXCHANGE	DE_NET_EXPORT	0.703962	0.703962
19	DE_FR_EXCHANGE	DE_NET_IMPORT	0.703962	-0.703962

6. Feature-Target Correlations

In [12]:
```
# Distribution of Target
plt.figure()
df["TARGET"].plot(kind="hist", bins=40, alpha=0.7)
plt.xlabel("TARGET")
plt.title(f"Distribution of TARGET")
plt.tight_layout()
plt.show()
```



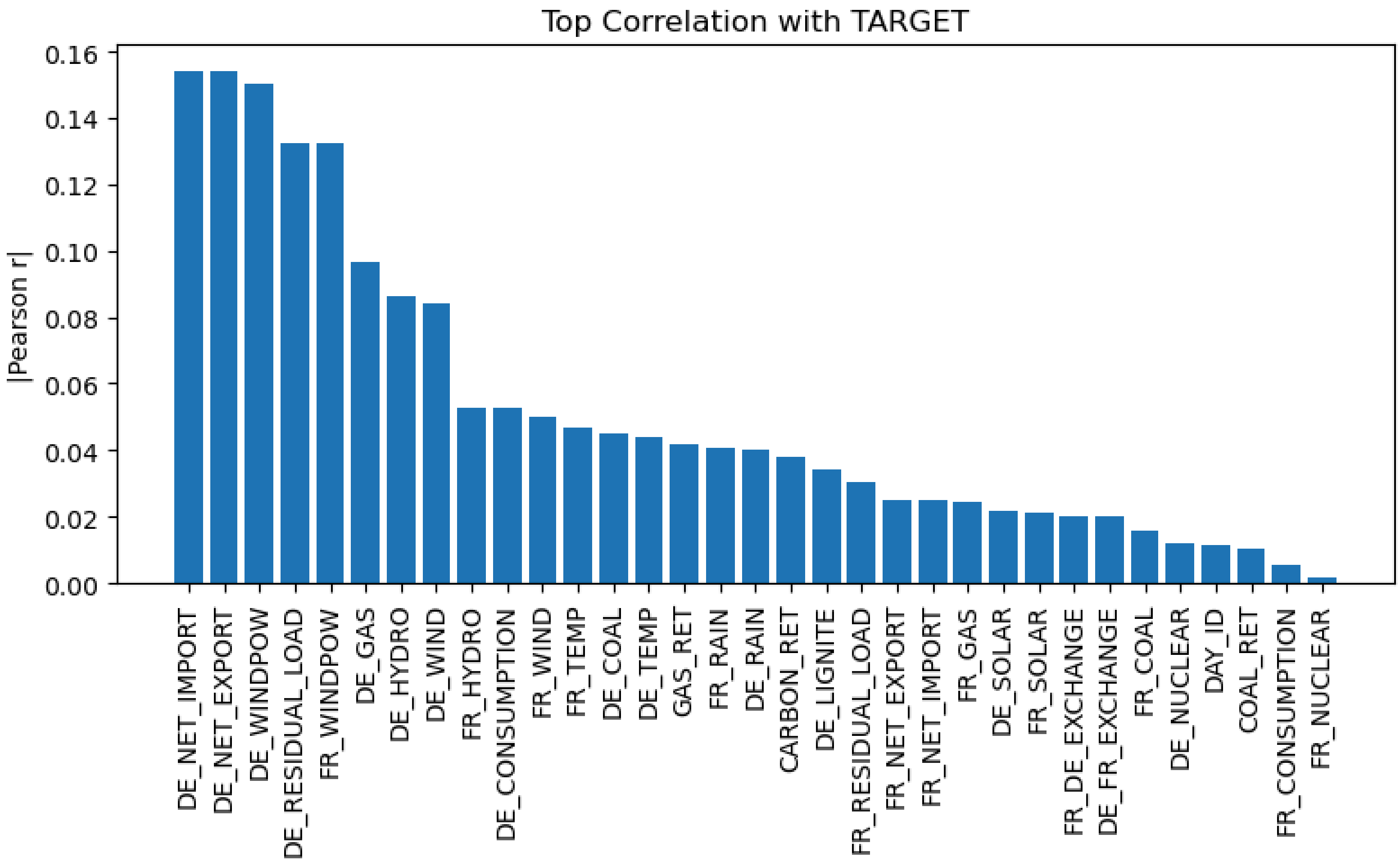
In [13]:
```
assert "TARGET" in df.columns, "TARGET missing after merge"
target_corr = df[num_cols + ["TARGET"]].corr()["TARGET"].drop("TARGET").sort_values(ascending=False)
```

```
display(target_corr)
```

```
DE_NET_IMPORT        0.154301
DE_RESIDUAL_LOAD     0.132555
DE_GAS               0.096657
DE_HYDRO             0.086127
FR_HYDRO             0.052820
DE_COAL              0.045181
GAS_RET              0.041969
CARBON_RET           0.038252
DE_LIGNITE           0.034433
FR_RESIDUAL_LOAD     0.030254
FR_NET_IMPORT        0.025213
FR_GAS               0.024652
DE_SOLAR             0.021934
FR_SOLAR             0.021081
FR_DE_EXCHANGE       0.020330
FR_COAL              0.015732
DE_NUCLEAR           0.012236
FR_NUCLEAR           0.001588
FR_CONSUMPTION      -0.005685
COAL_RET            -0.010285
DAY_ID              -0.011612
DE_FR_EXCHANGE      -0.020330
FR_NET_EXPORT       -0.025213
DE_RAIN             -0.040206
FR_RAIN             -0.040926
DE_TEMP             -0.044189
FR_TEMP             -0.046516
FR_WIND             -0.050216
DE_CONSUMPTION      -0.052565
DE_WIND             -0.084127
FR_WINDPOW          -0.132523
DE_WINDPOW          -0.150287
DE_NET_EXPORT       -0.154301
Name: TARGET, dtype: float64
```

In [14]:
```
# Bar of absolute correlations (top-k)
tc_abs = target_corr.abs().sort_values(ascending=False)
plt.figure()
plt.bar(tc_abs.index.astype(str), tc_abs.values)
```

```
plt.xticks(rotation=90)
plt.ylabel("|Pearson r|")
plt.title("Top Correlation with TARGET")
plt.tight_layout()
plt.show()
```



7) Time Series Analysis

In [15]:
```
def ensure_time_sorted(d):
    if "DAY_ID" in d.columns:
```

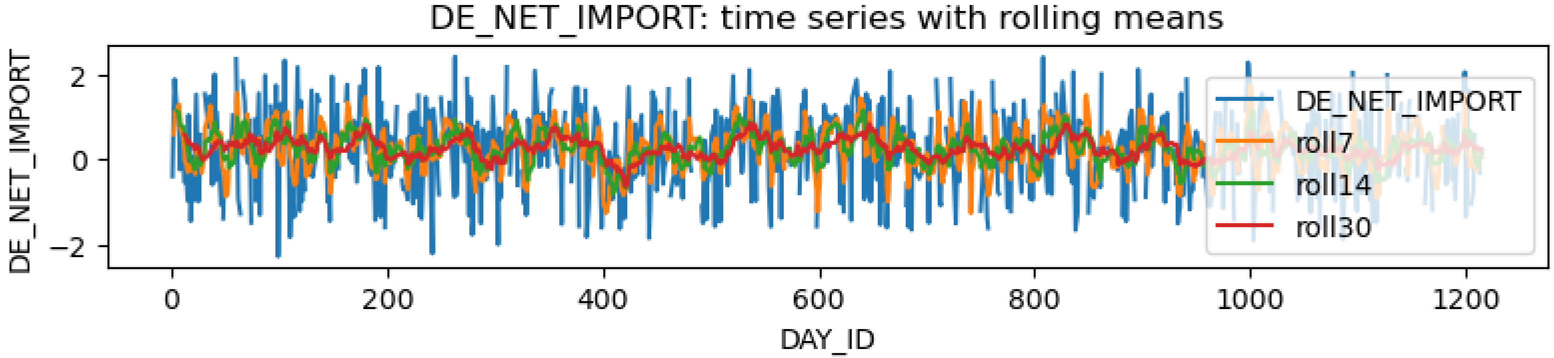
```
        return d.sort_values("DAY_ID")
    return d.copy()

def acf_values(x, max_lag=30):
    x = pd.Series(x).astype(float).to_numpy()
    x = x - np.nanmean(x)
    n = len(x)
    acf = []
    denom = np.nansum(x*x)
    for lag in range(1, max_lag+1):
        if lag >= n:
            acf.append(np.nan)
            continue
        a = x[:-lag]
        b = x[lag:]
        num = np.nansum(a*b)
        acf.append(num / denom if denom != 0 else np.nan)
    return np.array(acf)

def lagged_corr(x, y, max_lag=30):
    x = pd.Series(x).astype(float).to_numpy()
    y = pd.Series(y).astype(float).to_numpy()
    lags = list(range(-max_lag, max_lag+1))
    vals = []
    for L in lags:
        if L < 0:
            xx = x[-L:]; yy = y[:len(y)+L]
        elif L > 0:
            xx = x[:-L]; yy = y[L:]
        else:
            xx = x; yy = y
        if len(xx) < 3:
            vals.append(np.nan); continue
        mask = np.isfinite(xx) & np.isfinite(yy)
        if mask.sum() < 3:
            vals.append(np.nan); continue
        xv = xx[mask]; yv = yy[mask]
        if np.std(xv)==0 or np.std(yv)==0:
            vals.append(np.nan); continue
```

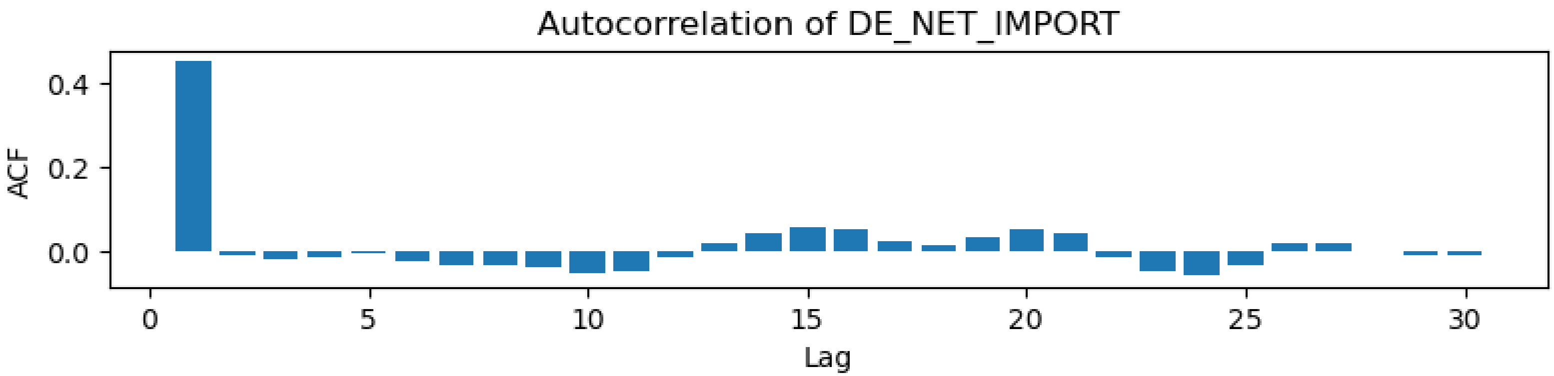
```
        vals.append(np.corrcoef(xv, yv)[0,1])
    return np.array(lags), np.array(vals)
```

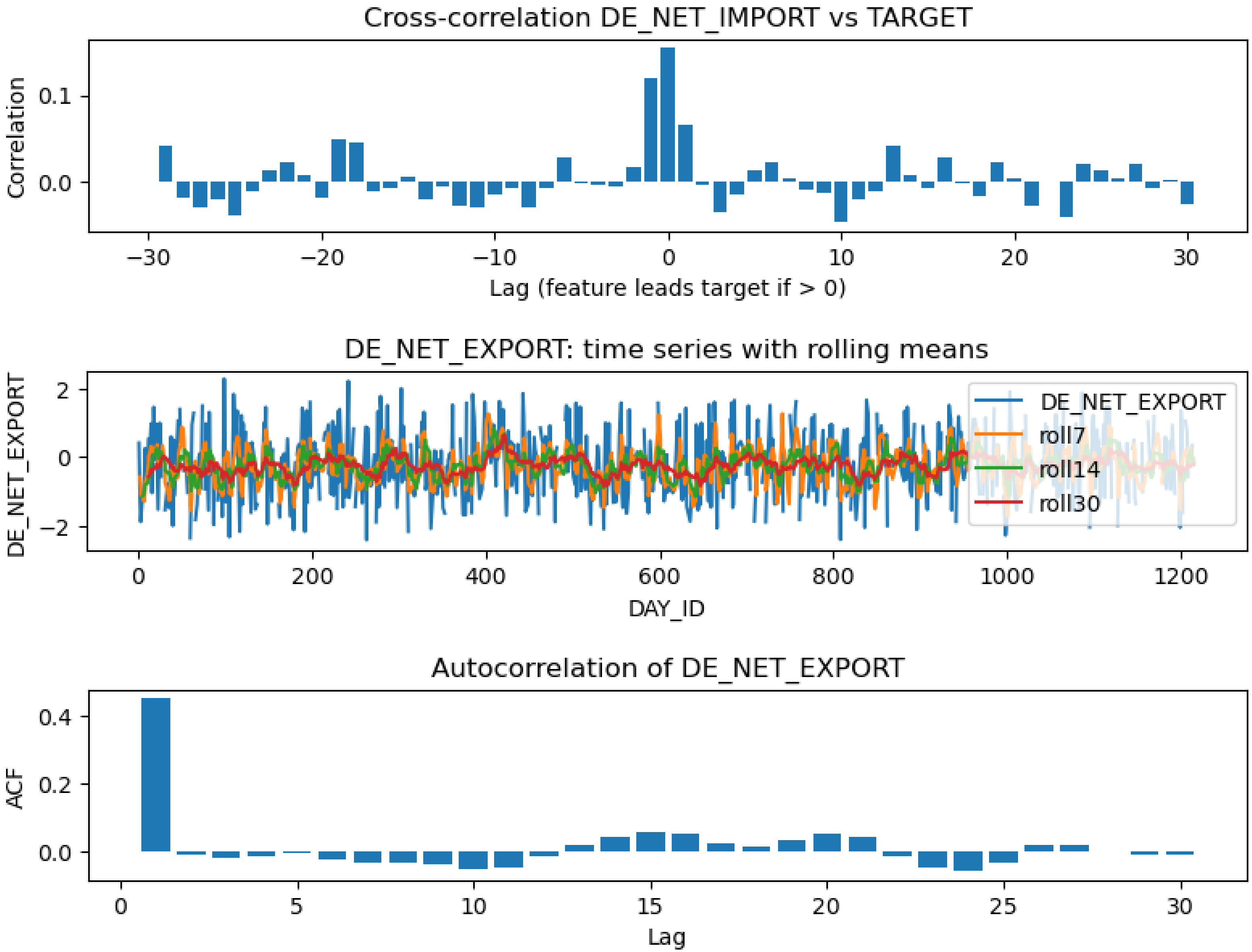
In [16]:
```
if "DAY_ID" not in df.columns:
    print("DAY_ID not found; skipping TS analysis.")
else:
    d = ensure_time_sorted(df)
    ts_candidates = d[num_cols[1:] + ["TARGET"]].corr()["TARGET"].drop("TARGET").abs().sort_values(ascending=False)
    ts_features = list(ts_candidates.index)
    print("TS features (global):", ts_features)
    yv = d["TARGET"].astype(float).to_numpy()

    for c in ts_features:
        # Original + rolling means
        s = d[c].astype(float)
        plt.figure()
        plt.subplot(311)
        plt.plot(d["DAY_ID"], s, label=f"{c}")
        for w in ROLL_WINDOWS:
            try:
                plt.plot(d["DAY_ID"], s.rolling(window=w, min_periods=max(1,w//2)).mean(), label=f"roll{w}")
            except Exception:
                pass
        plt.xlabel("DAY_ID"); plt.ylabel(c)
        plt.title(f"{c}: time series with rolling means")
        plt.legend()
        plt.tight_layout()
        plt.show()

        # ACF
        vals = d[c].astype(float).to_numpy()
        acf = acf_values(vals)
        plt.subplot(312)
        plt.bar(np.arange(1, len(acf)+1), acf)
        plt.xlabel("Lag"); plt.ylabel("ACF")
        plt.title(f"Autocorrelation of {c}")
        plt.tight_layout()
        plt.show()

        # Cross-correlation with TARGET
```

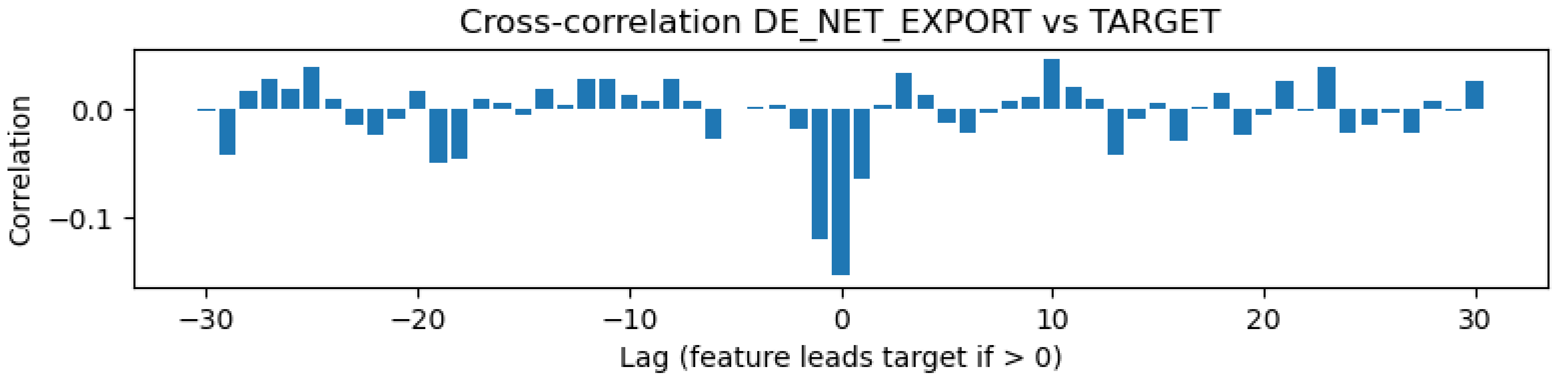
```
        xv = d[c].astype(float).to_numpy()
        lags, cc = lagged_corr(xv, yv)
        plt.subplot(313)
        plt.bar(lags, cc)
        plt.xlabel("Lag (feature leads target if > 0)")
        plt.ylabel("Correlation")
        plt.title(f"Cross-correlation {c} vs TARGET")
        plt.tight_layout()
        plt.show()
```

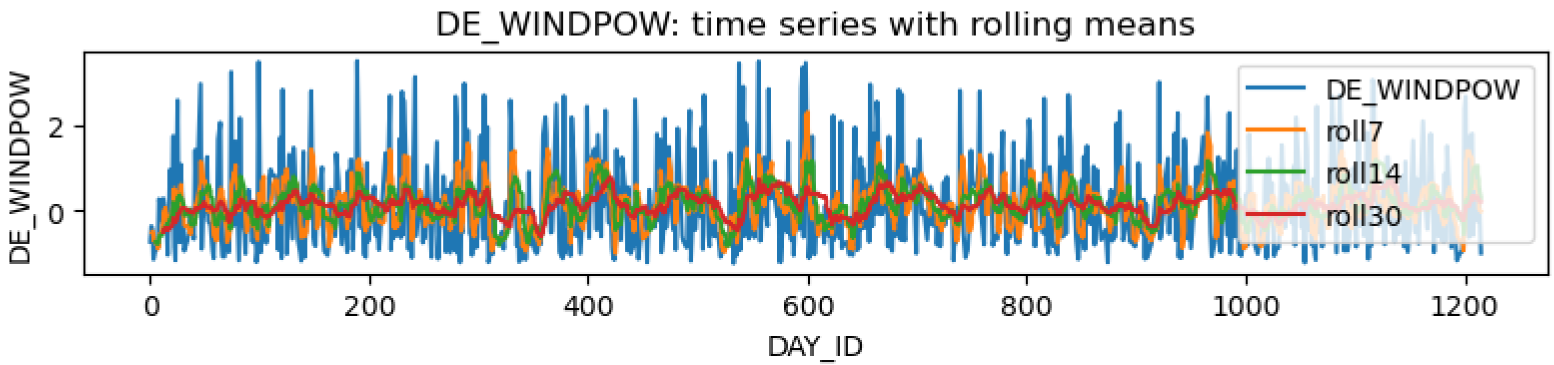
```
TS features (global): ['DE_NET_IMPORT', 'DE_NET_EXPORT', 'DE_WINDPOW', 'DE_RESIDUAL_LOAD', 'FR_WINDPOW', 'DE_GAS', 'DE_HYDRO',
'DE_WIND', 'FR_HYDRO', 'DE_CONSUMPTION', 'FR_WIND', 'FR_TEMP', 'DE_COAL', 'DE_TEMP', 'GAS_RET', 'FR_RAIN', 'DE_RAIN', 'CARBON_R
ET', 'DE_LIGNITE', 'FR_RESIDUAL_LOAD', 'FR_NET_EXPORT', 'FR_NET_IMPORT', 'FR_GAS', 'DE_SOLAR', 'FR_SOLAR', 'DE_FR_EXCHANGE', 'F
R_DE_EXCHANGE', 'FR_COAL', 'DE_NUCLEAR', 'COAL_RET', 'FR_CONSUMPTION', 'FR_NUCLEAR']
```

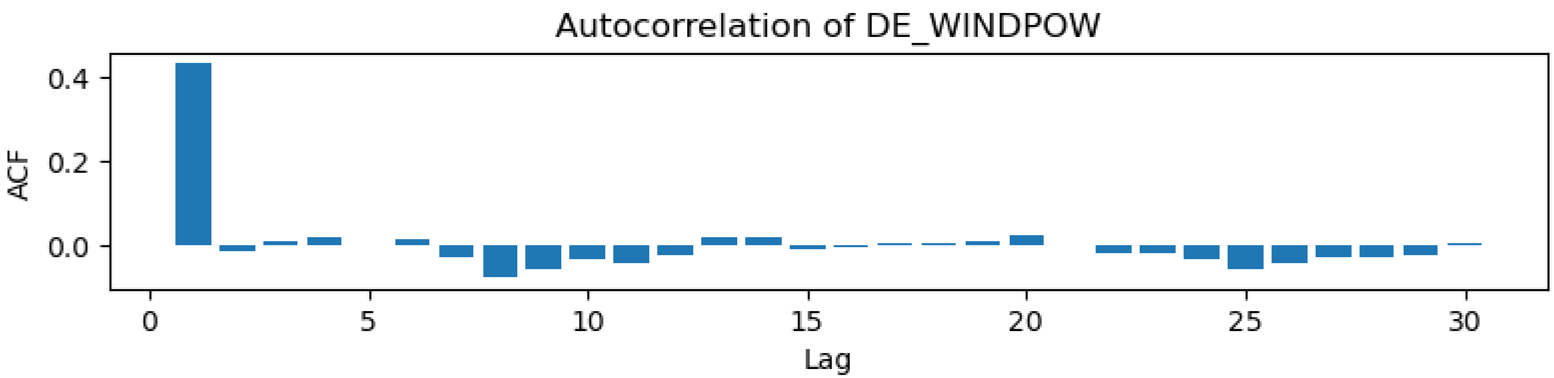


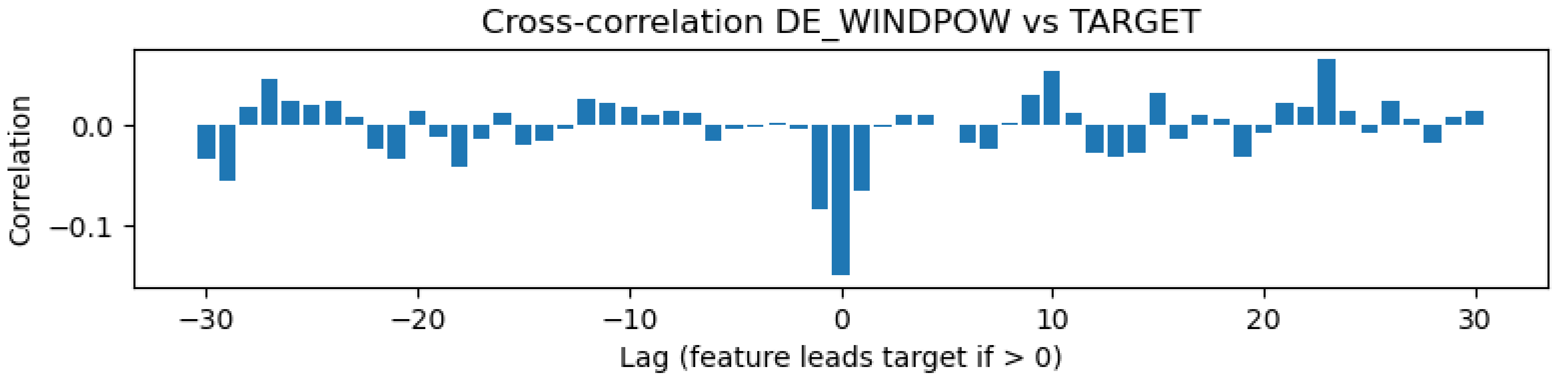


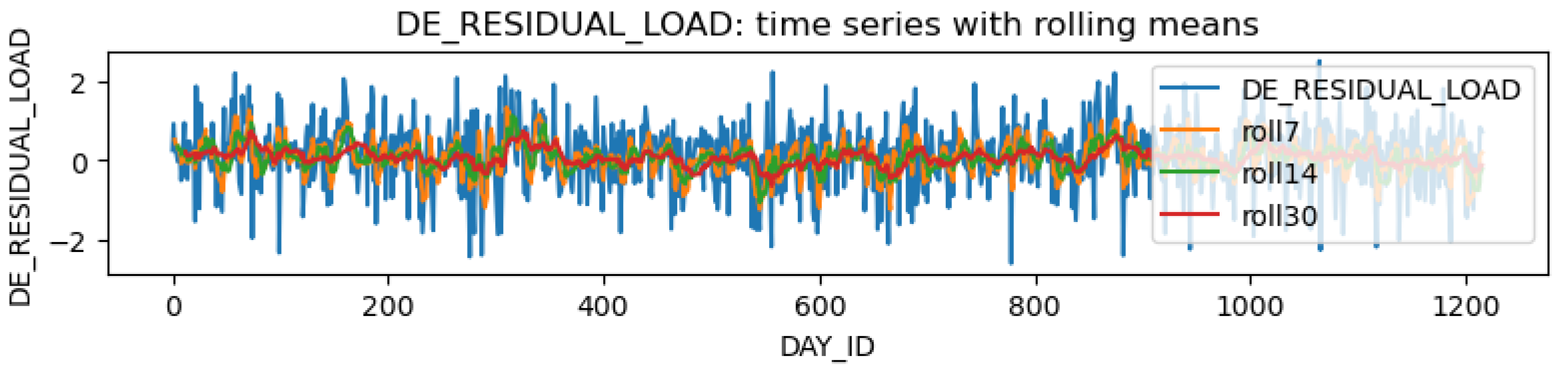

Cross-correlation DE_NET_IMPORT vs TARGET
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)
DE_NET_EXPORT: time series with rolling means
DE_NET_EXPORT
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_NET_EXPORT
roll7
roll14
roll30
Autocorrelation of DE_NET_EXPORT
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

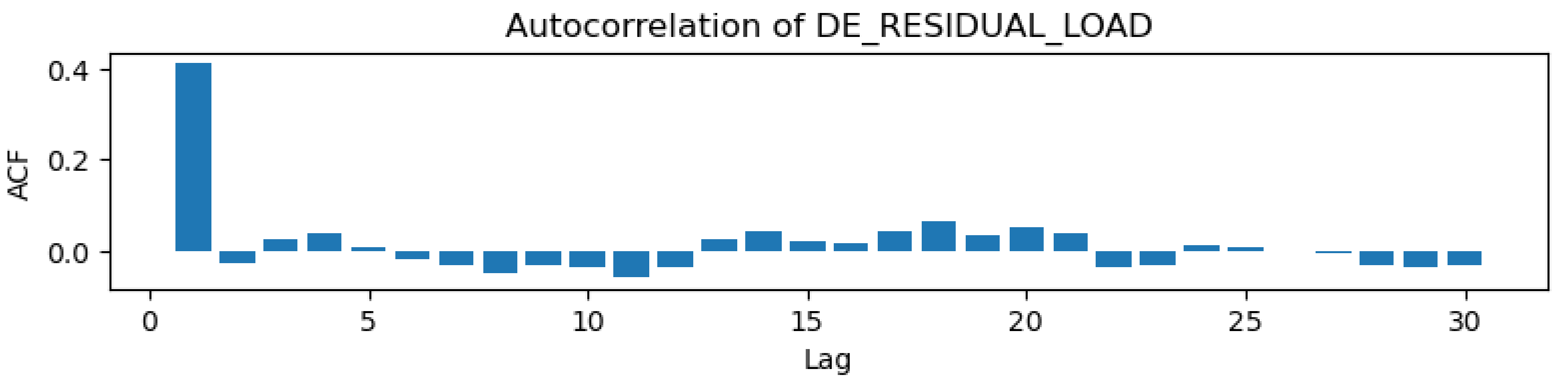

Cross-correlation DE_NET_EXPORT vs TARGET
Correlation
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

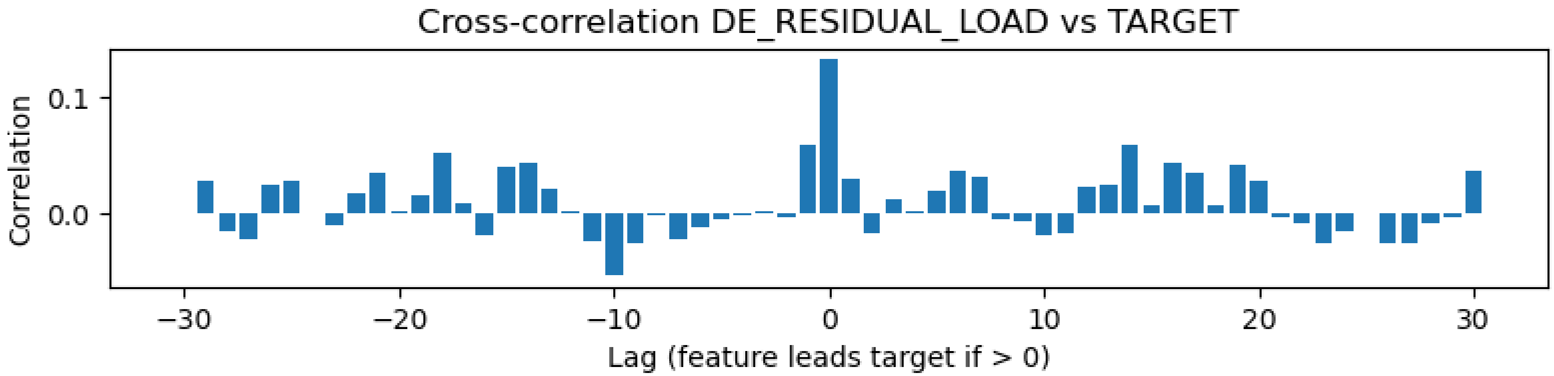

DE_WINDPOW: time series with rolling means
DE_WINDPOW
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_WINDPOW
roll7
roll14
roll30

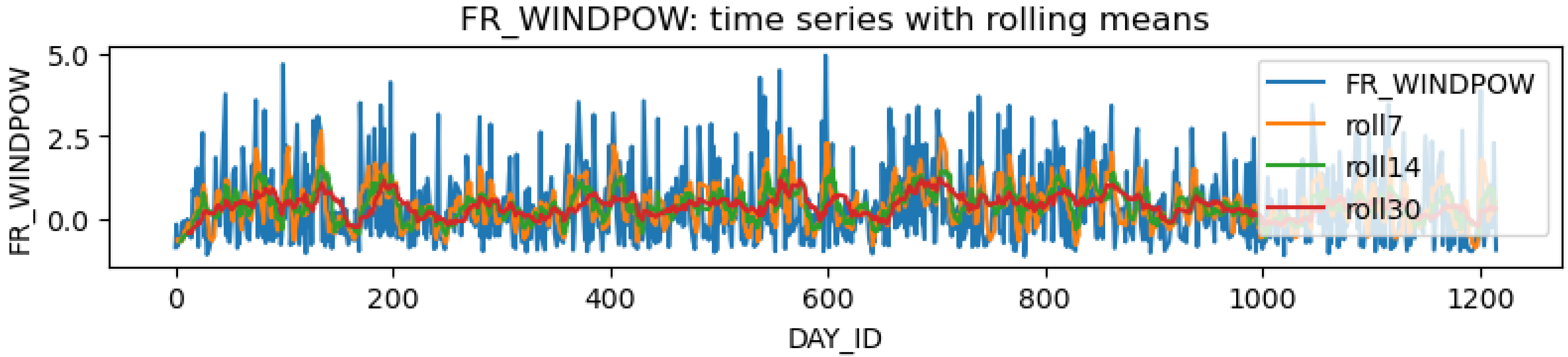

Autocorrelation of DE_WINDPOW
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

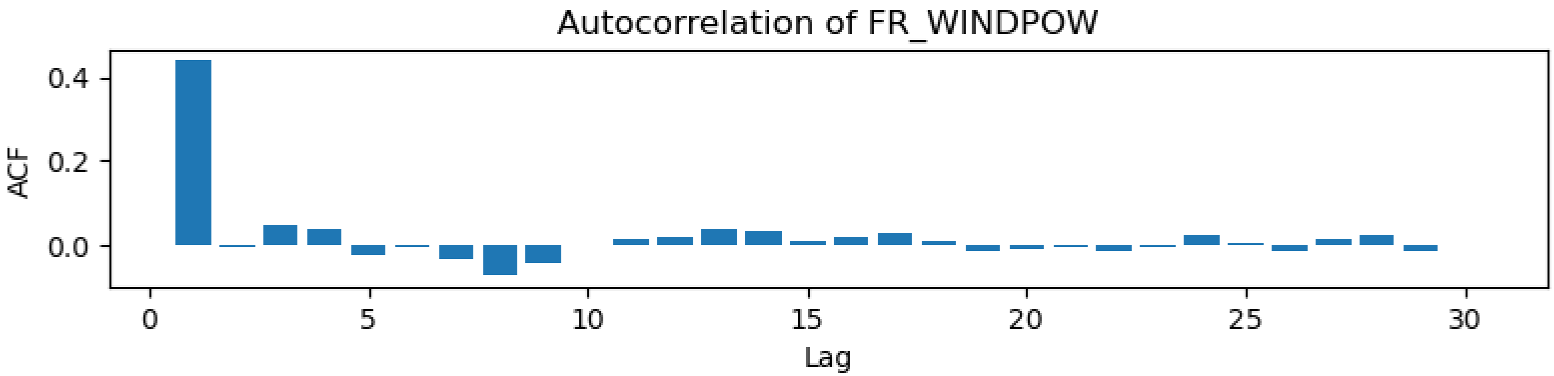

Cross-correlation DE_WINDPOW vs TARGET
Correlation
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

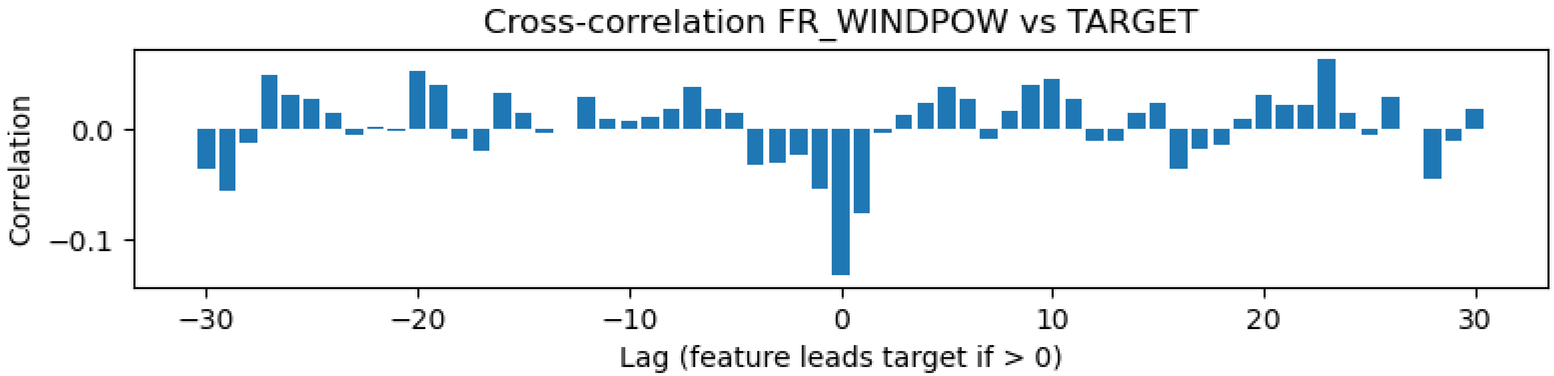

DE_RESIDUAL_LOAD: time series with rolling means
DE_RESIDUAL_LOAD
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_RESIDUAL_LOAD
roll7
roll14
roll30

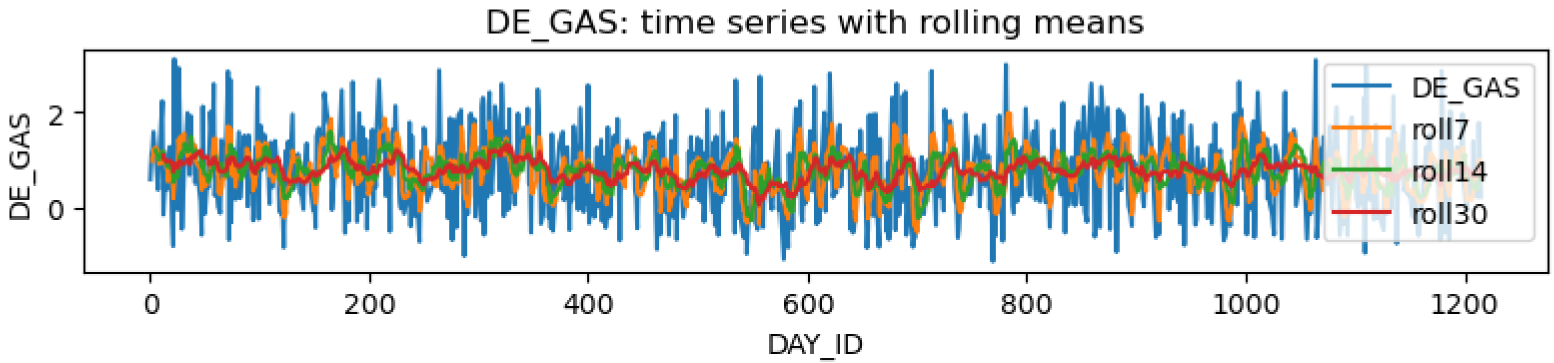

Autocorrelation of DE_RESIDUAL_LOAD
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

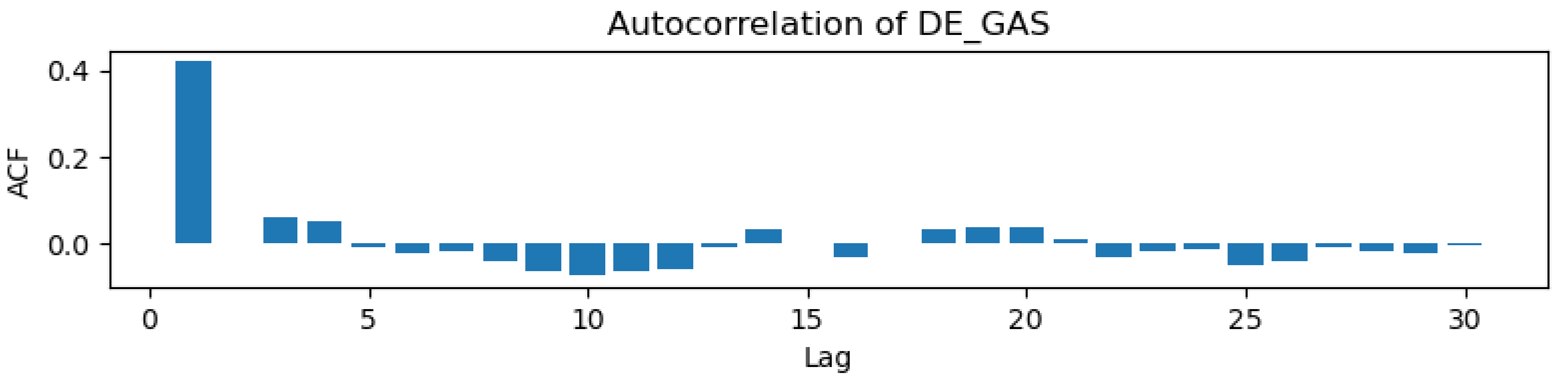

Cross-correlation DE_RESIDUAL_LOAD vs TARGET
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

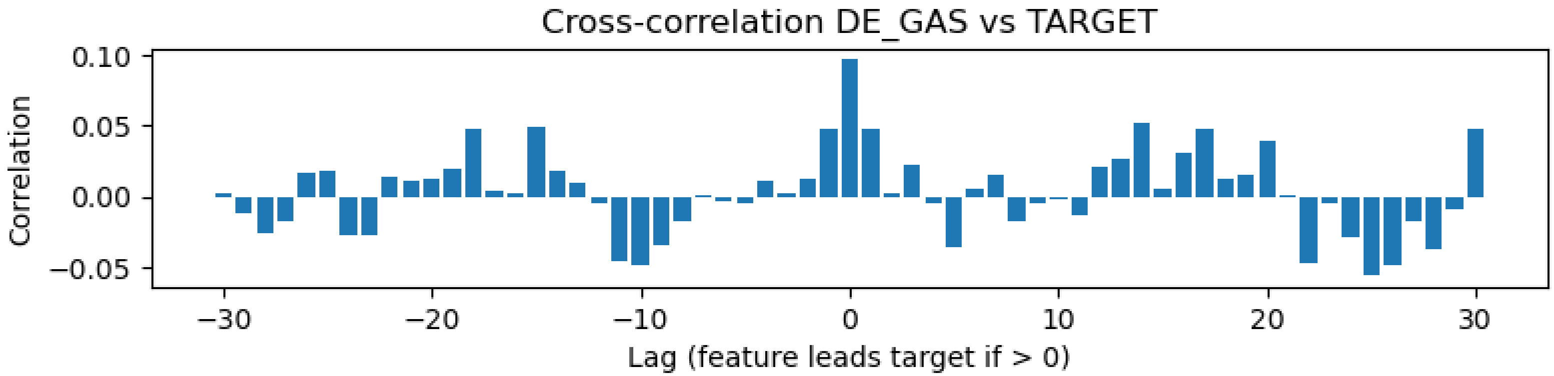

FR_WINDPOW: time series with rolling means
FR_WINDPOW
5.0
2.5
0.0
0
200
400
600
800
1000
1200
DAY_ID
FR_WINDPOW
roll7
roll14
roll30

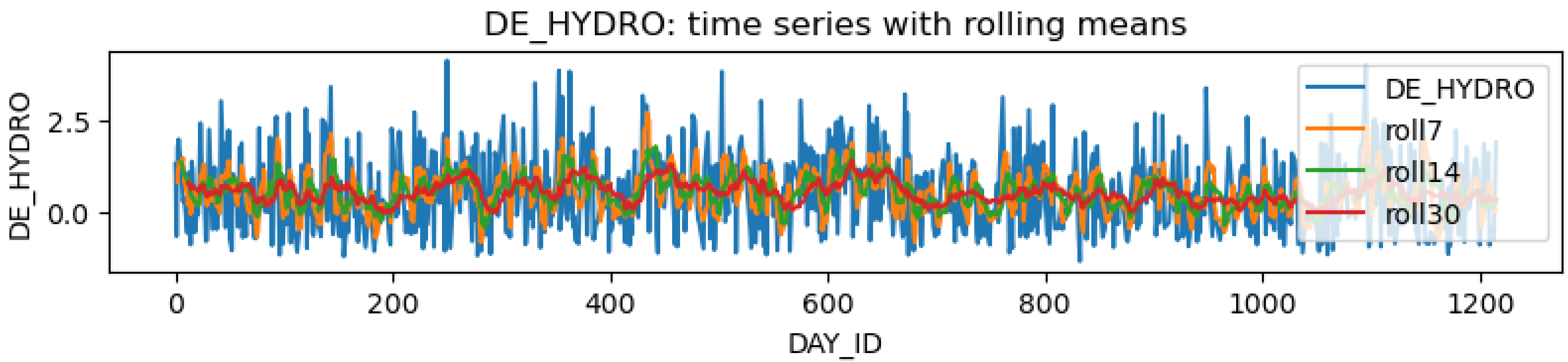

Autocorrelation of FR_WINDPOW
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

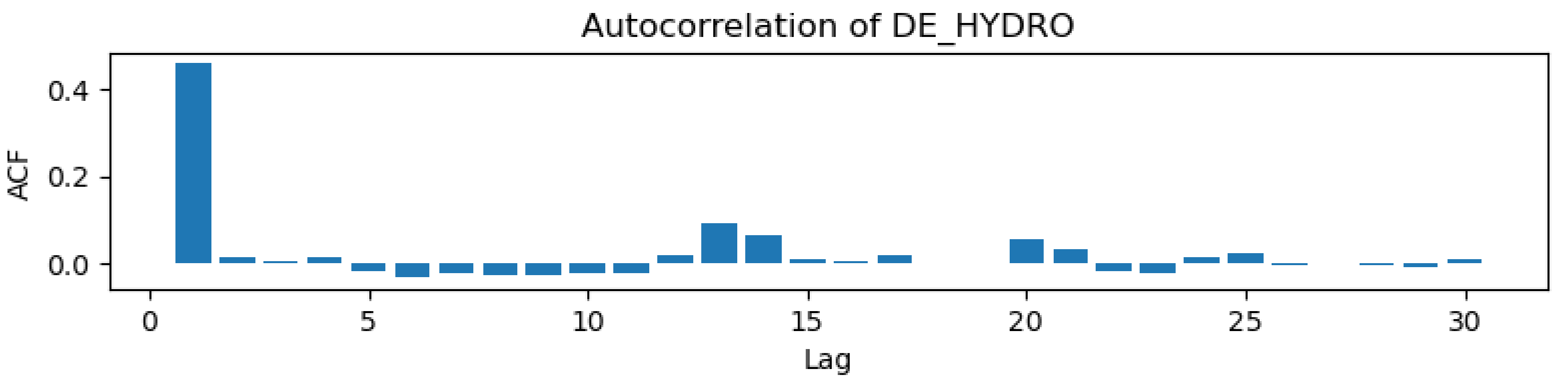

Cross-correlation FR_WINDPOW vs TARGET
Correlation
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

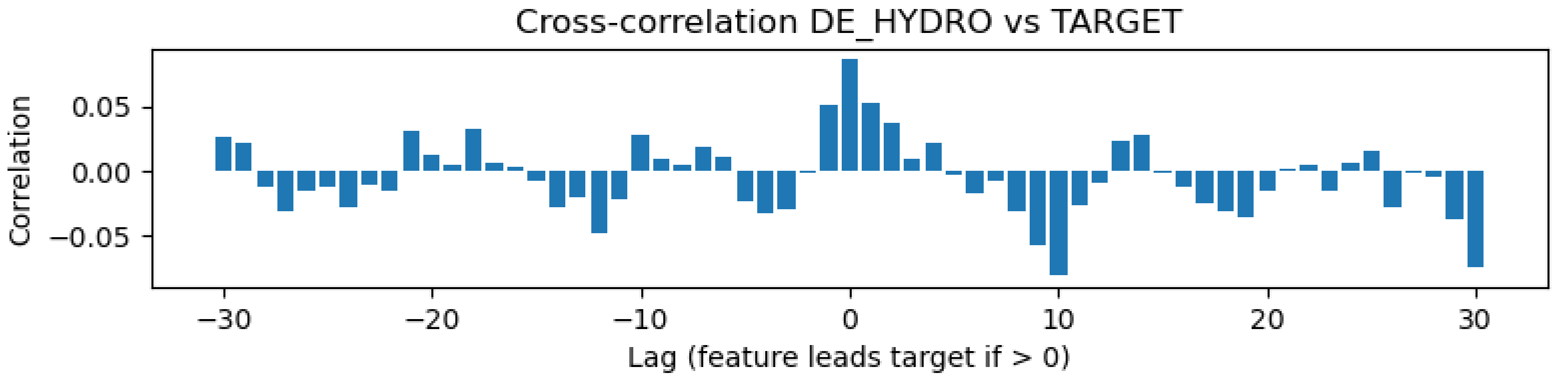

DE_GAS: time series with rolling means
DE_GAS
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_GAS
roll7
roll14
roll30

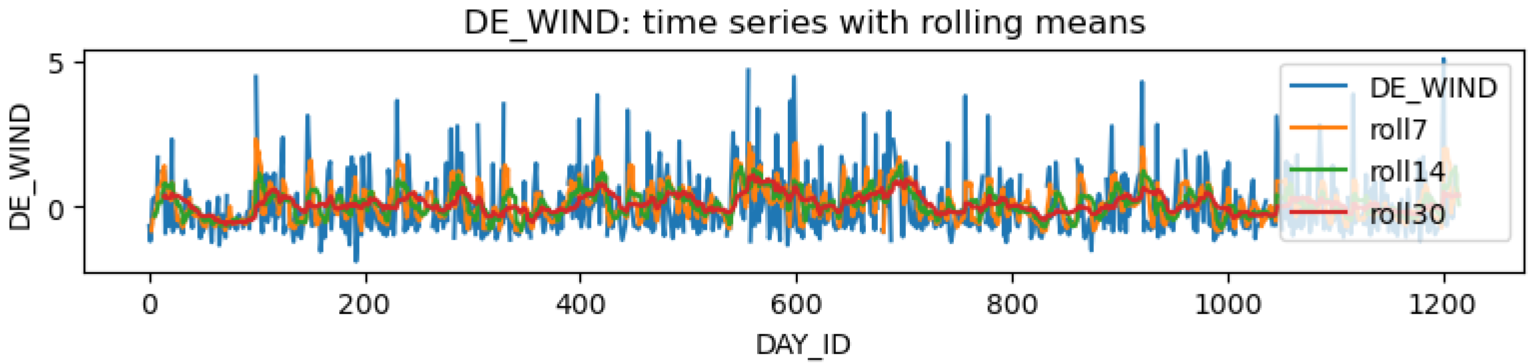

Autocorrelation of DE_GAS
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

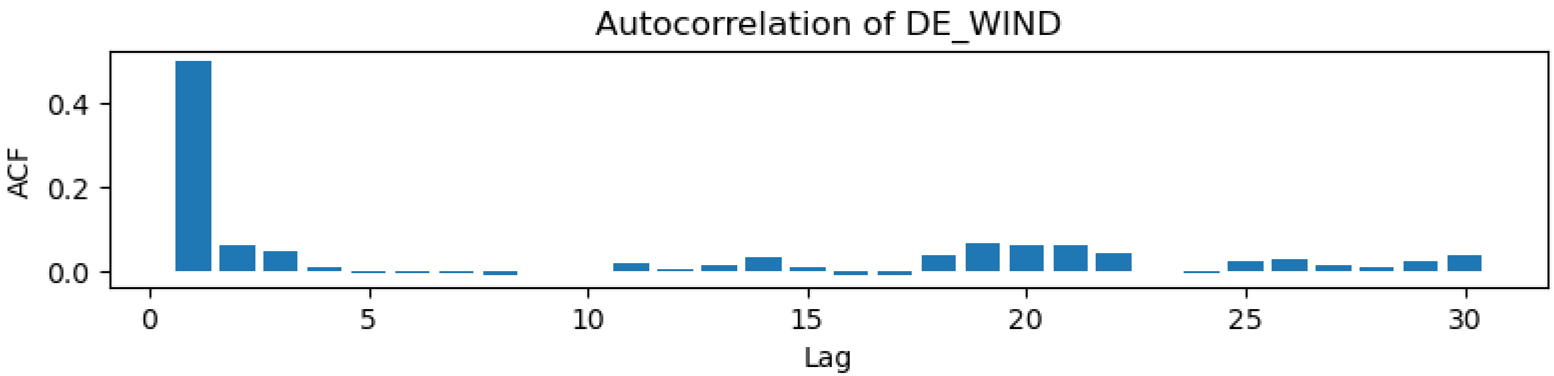

Cross-correlation DE_GAS vs TARGET
Correlation
0.10
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

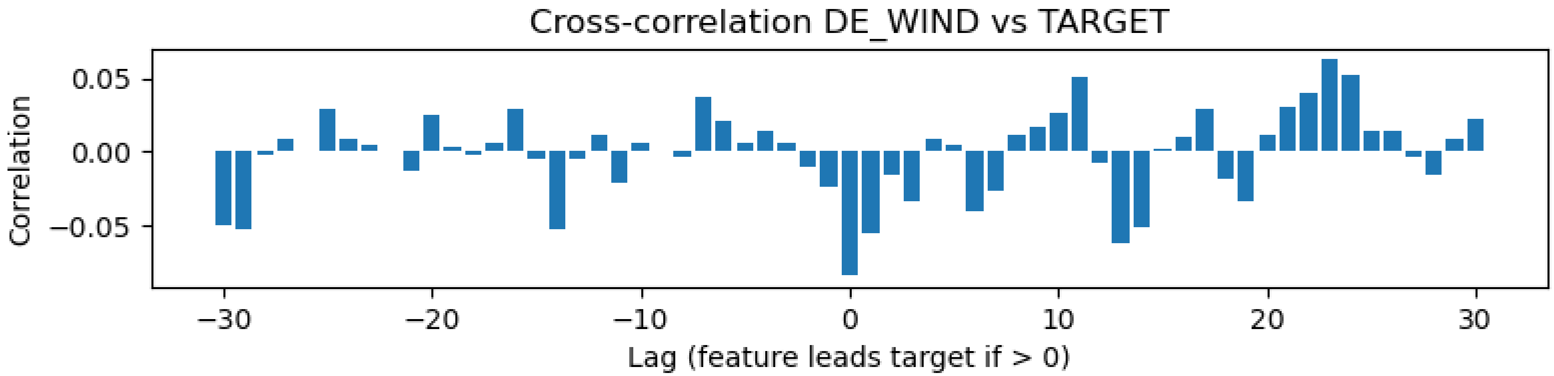

DE_HYDRO: time series with rolling means
DE_HYDRO
2.5
0.0
0
200
400
600
800
1000
1200
DAY_ID
DE_HYDRO
roll7
roll14
roll30

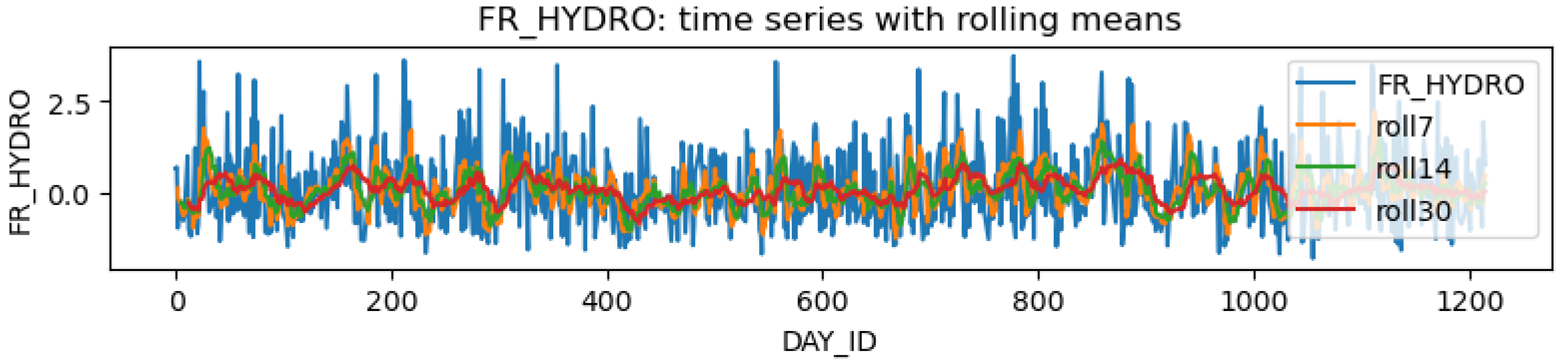

Autocorrelation of DE_HYDRO
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

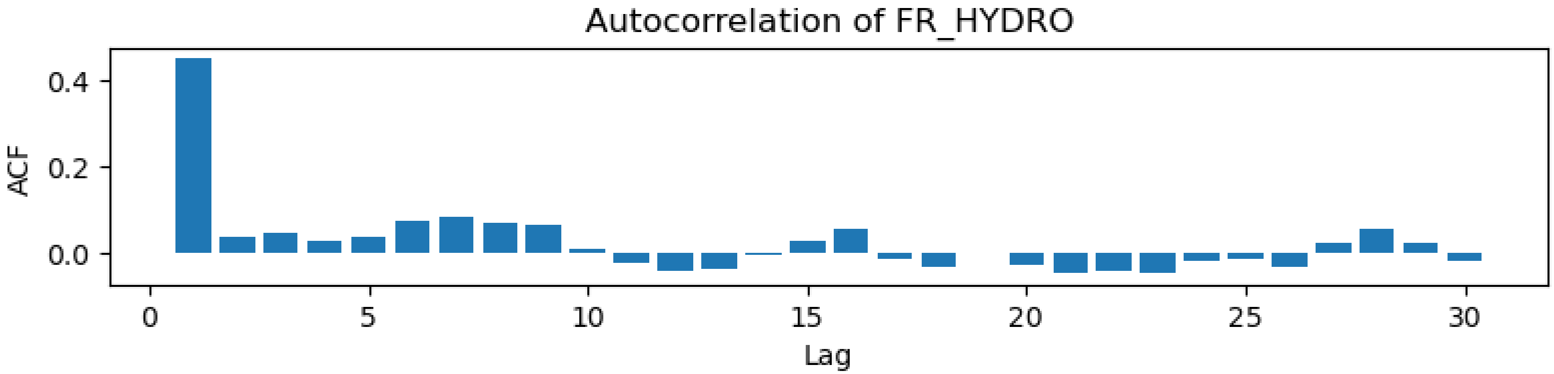

Cross-correlation DE_HYDRO vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

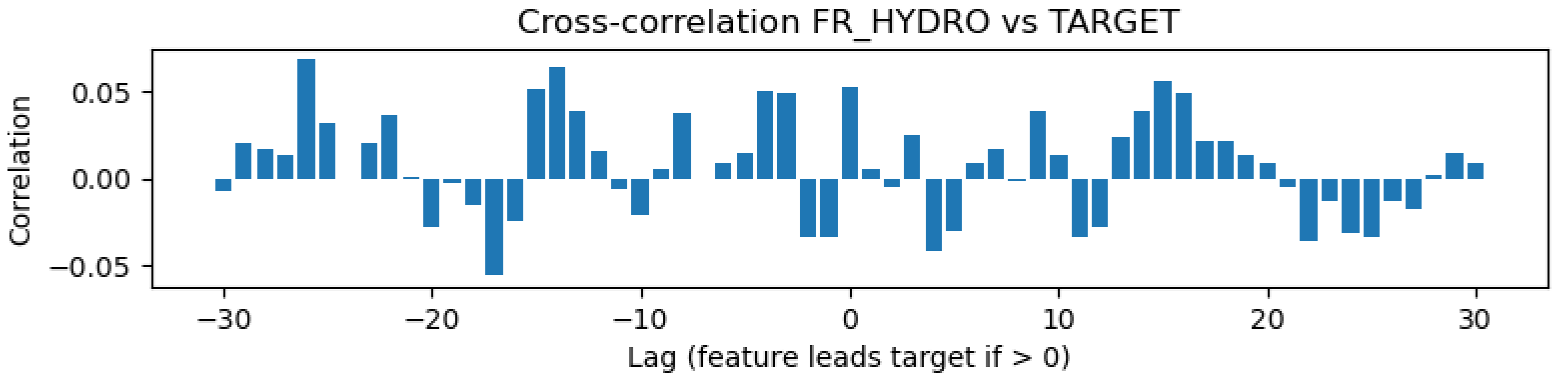

DE_WIND: time series with rolling means
DE_WIND
5
0
0
200
400
600
800
1000
1200
DAY_ID
DE_WIND
roll7
roll14
roll30

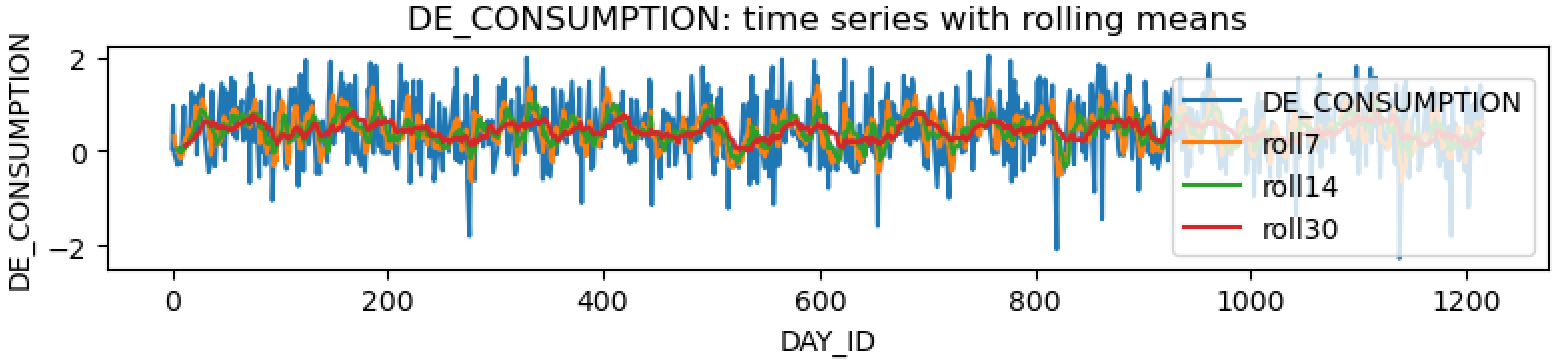

Autocorrelation of DE_WIND
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

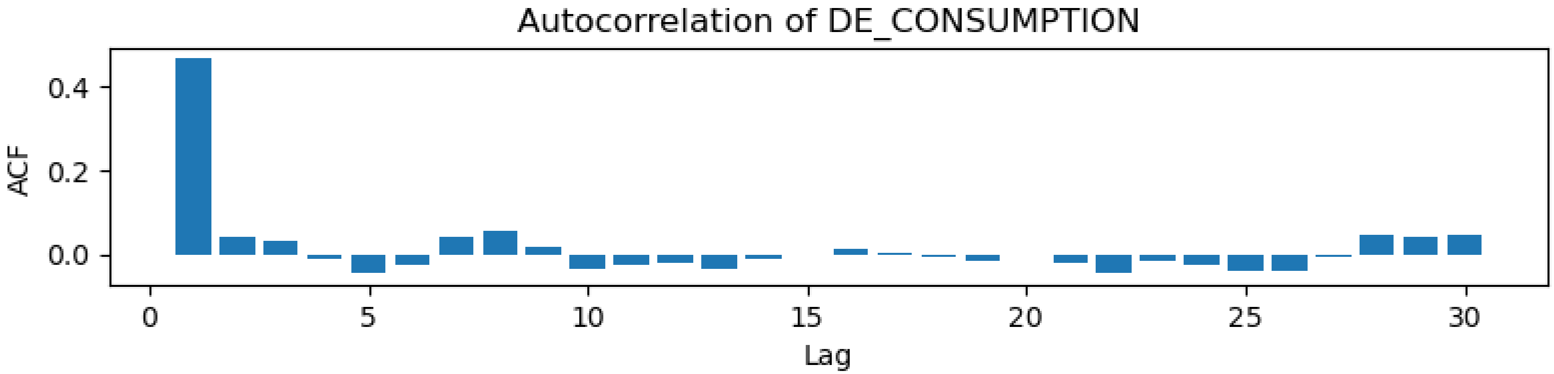

Cross-correlation DE_WIND vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

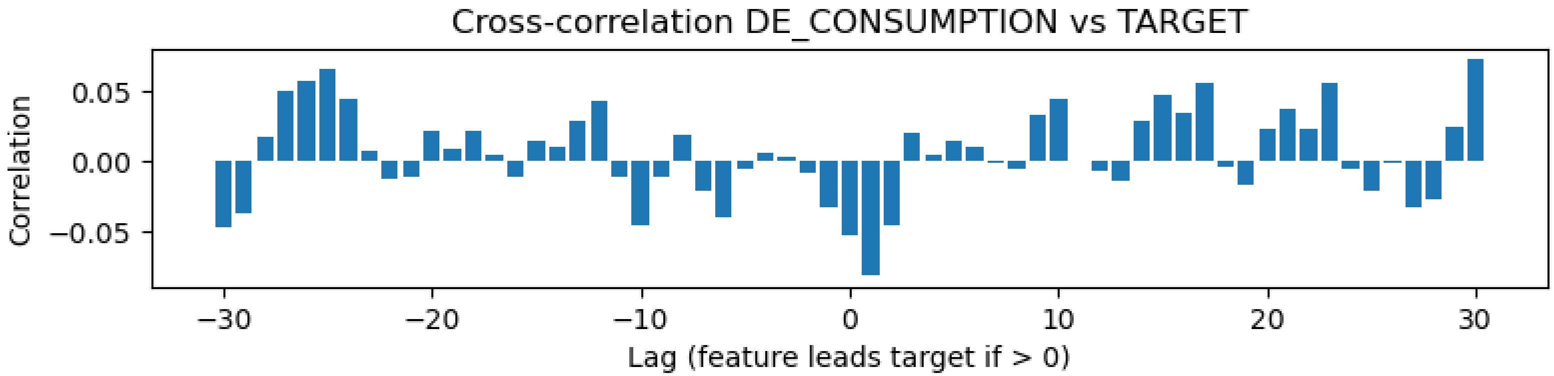

FR_HYDRO: time series with rolling means
FR_HYDRO
2.5
0.0
0
200
400
600
800
1000
1200
DAY_ID
FR_HYDRO
roll7
roll14
roll30

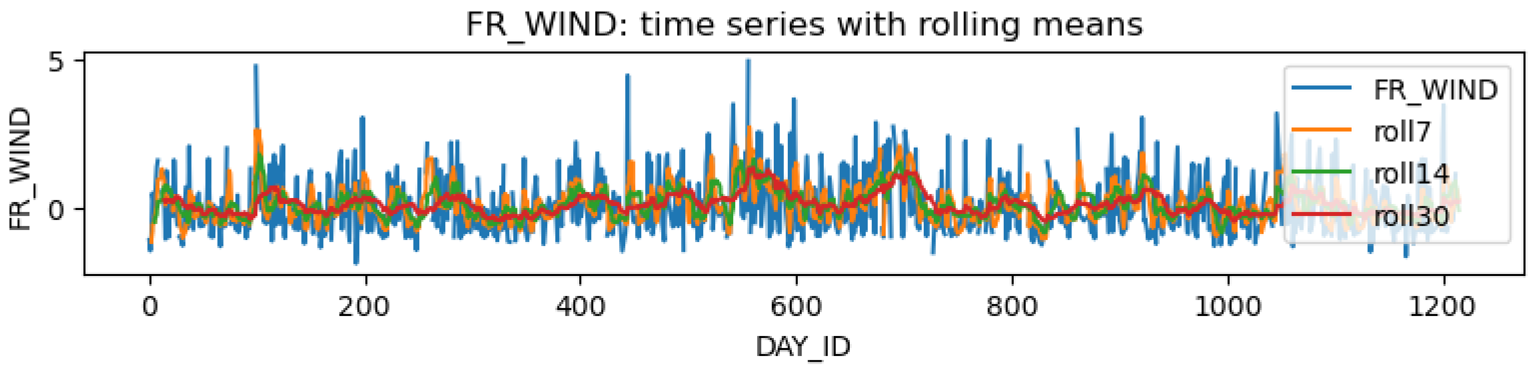

Autocorrelation of FR_HYDRO
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

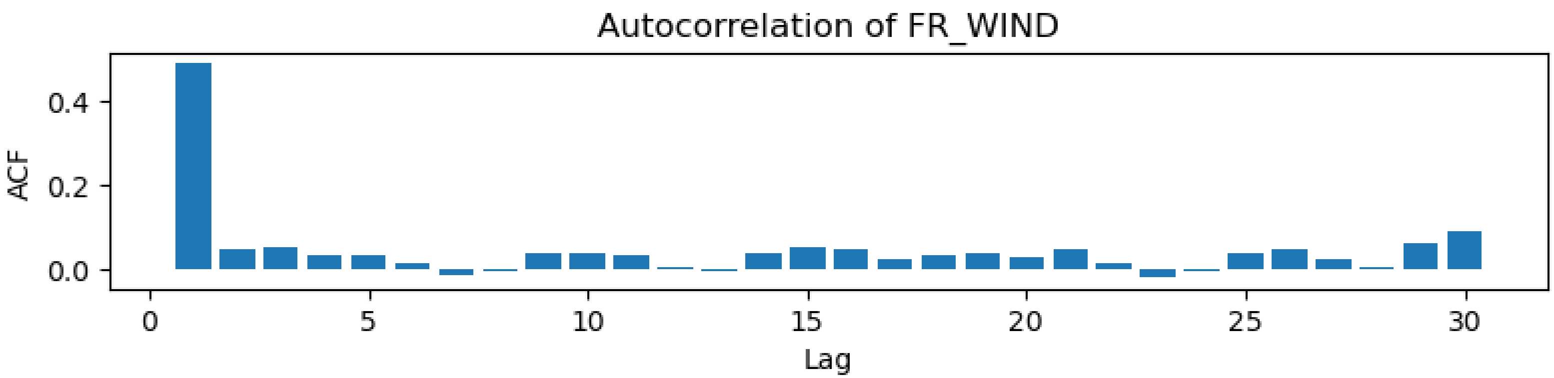

Cross-correlation FR_HYDRO vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

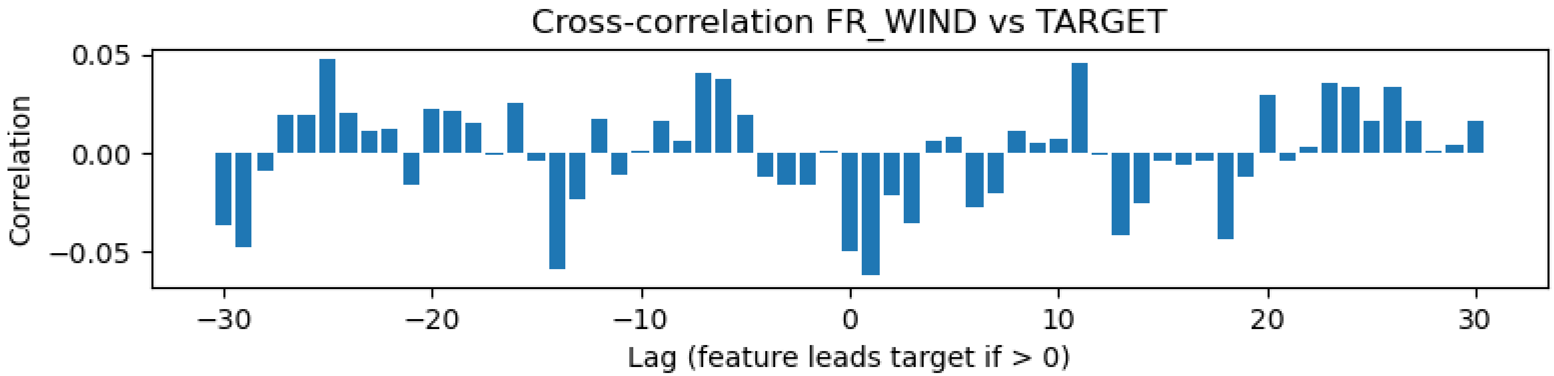

DE_CONSUMPTION: time series with rolling means
DE_CONSUMPTION
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_CONSUMPTION
roll7
roll14
roll30

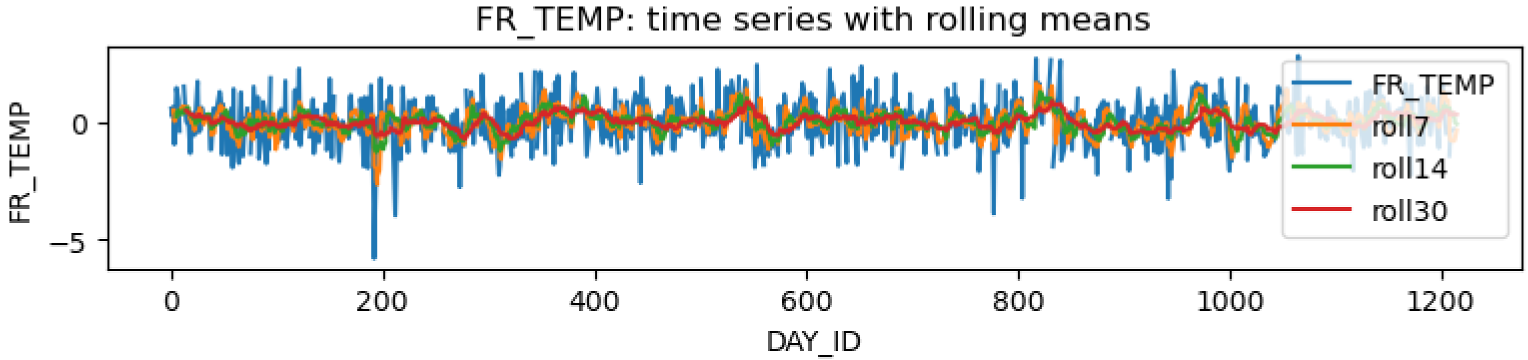

Autocorrelation of DE_CONSUMPTION
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

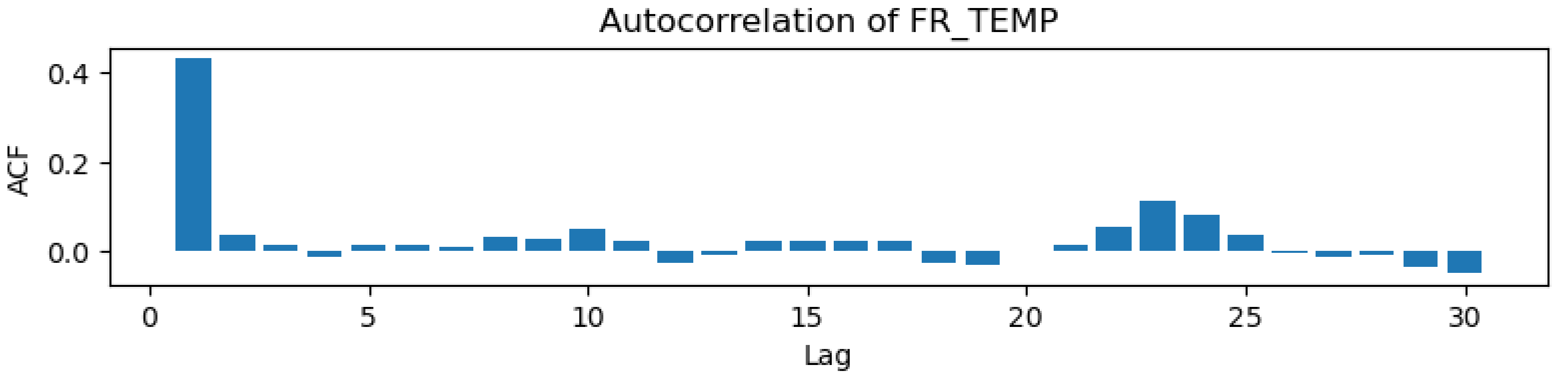

Cross-correlation DE_CONSUMPTION vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

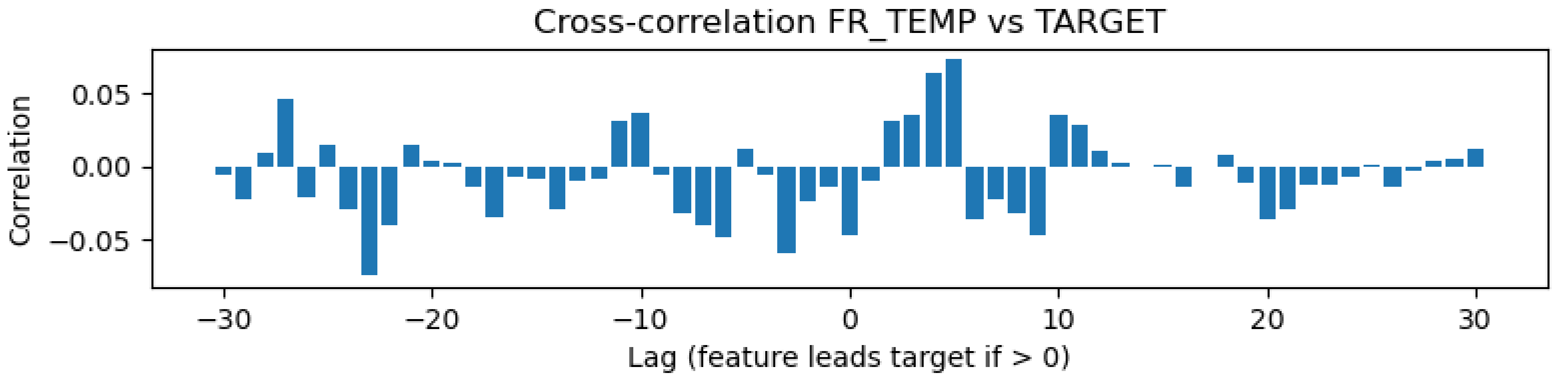

FR_WIND: time series with rolling means
FR_WIND
5
0
FR_WIND
roll7
roll14
roll30
0
200
400
600
800
1000
1200
DAY_ID

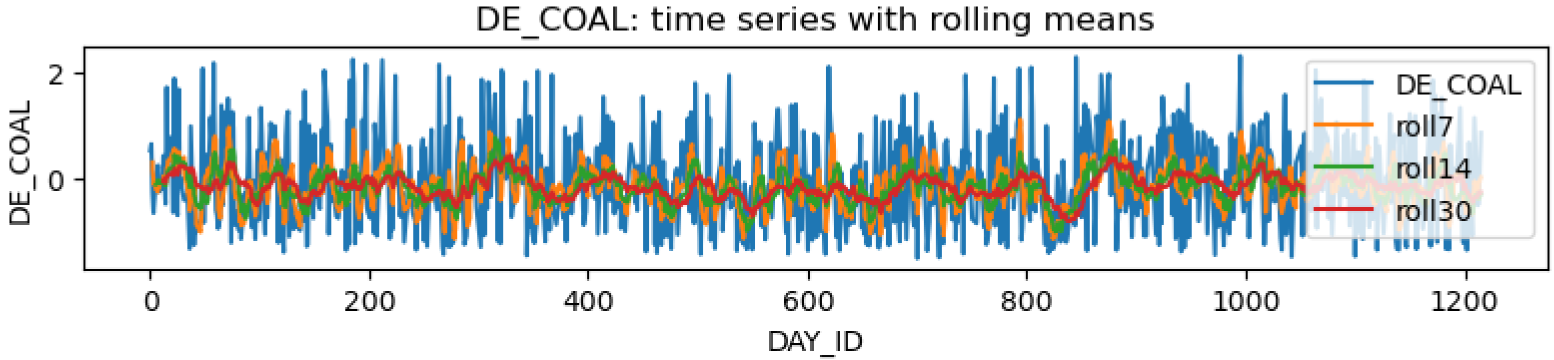

Autocorrelation of FR_WIND
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

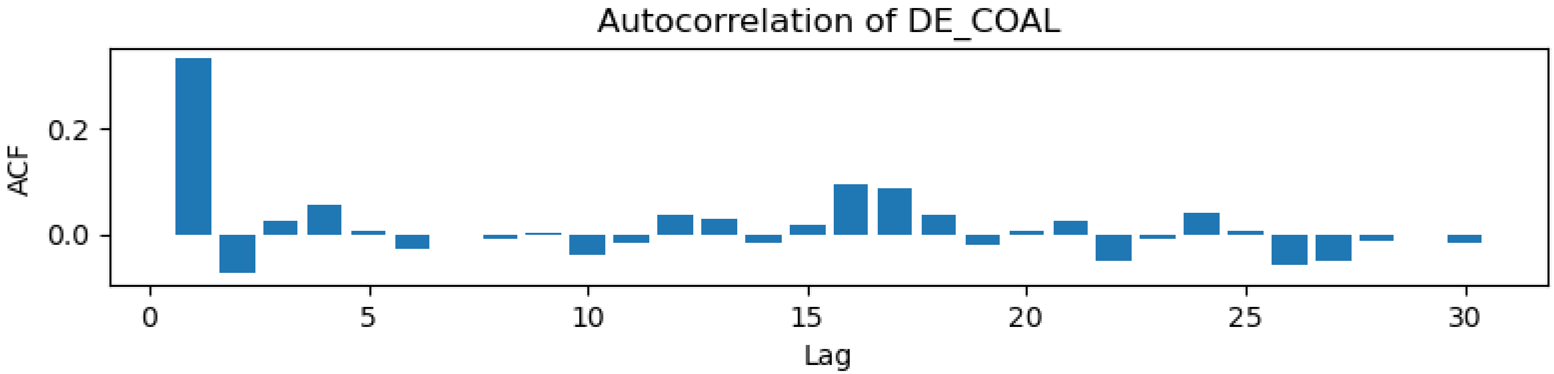

Cross-correlation FR_WIND vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

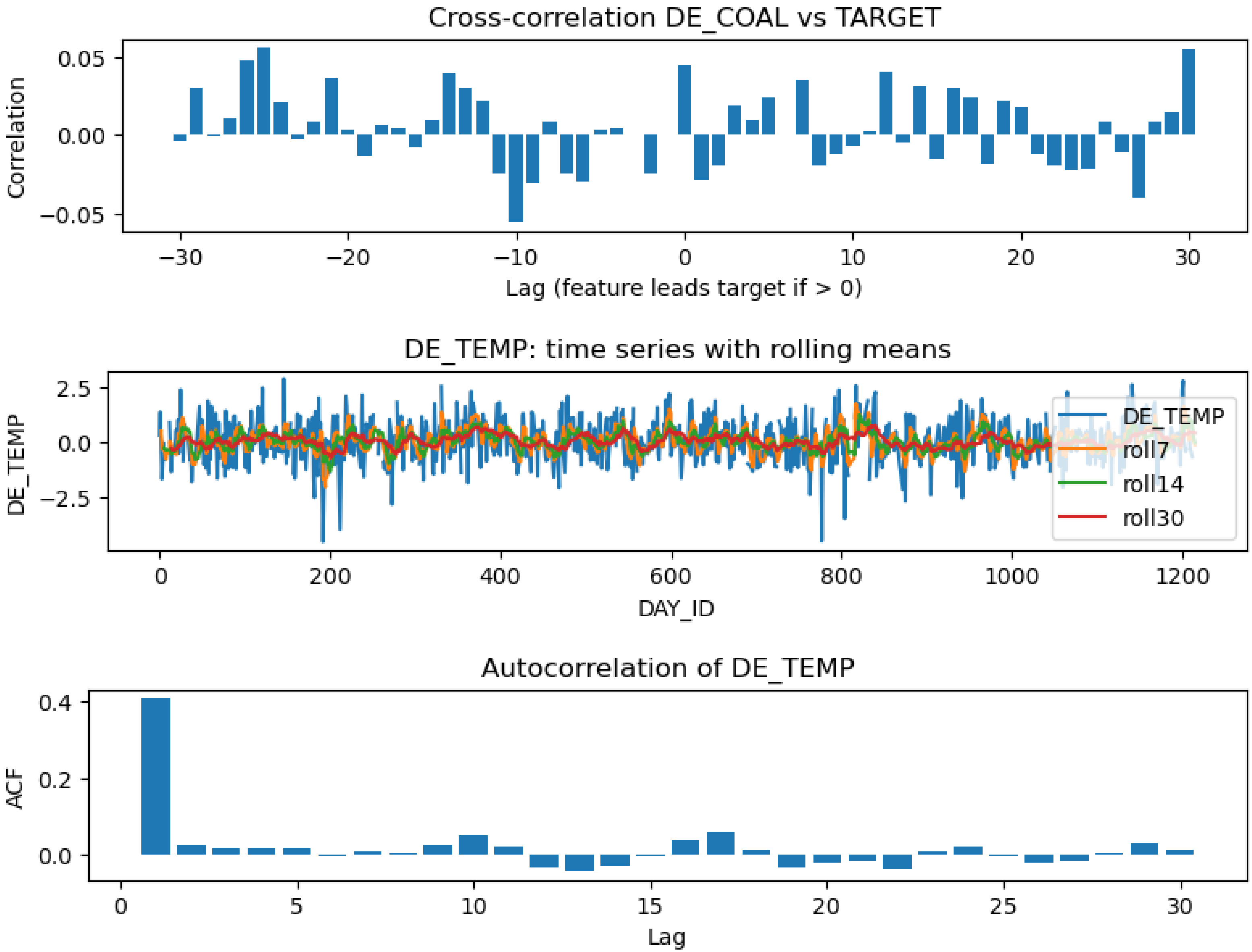

FR_TEMP: time series with rolling means
FR_TEMP
0
−5
0
200
400
600
800
1000
1200
DAY_ID
FR_TEMP
roll7
roll14
roll30

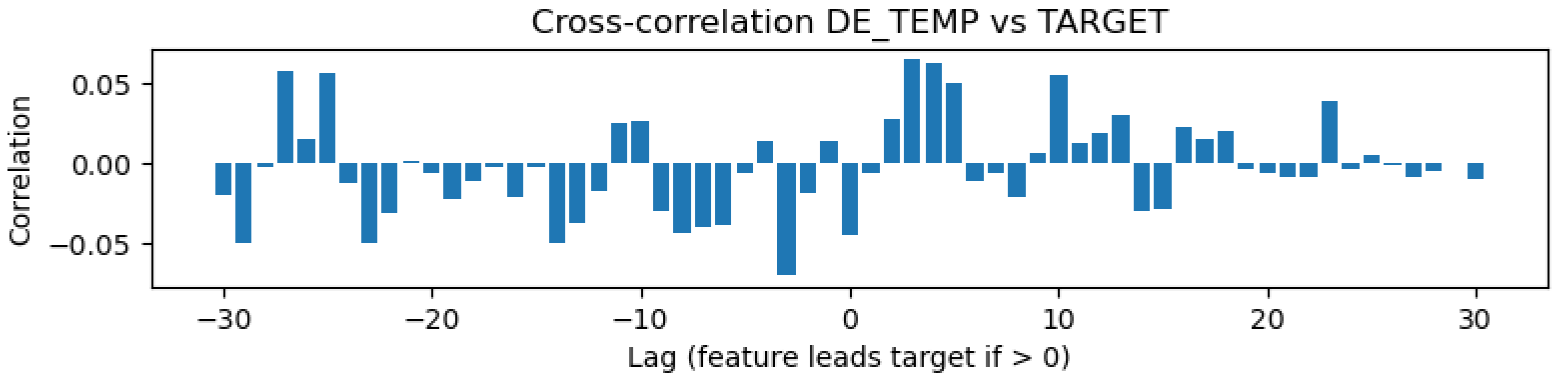

Autocorrelation of FR_TEMP
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

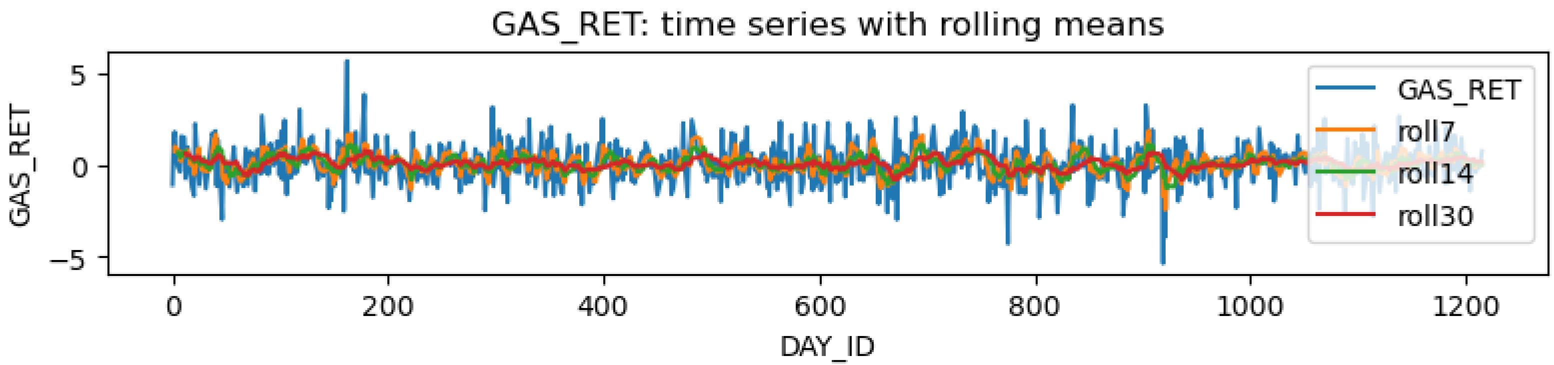

Cross-correlation FR_TEMP vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

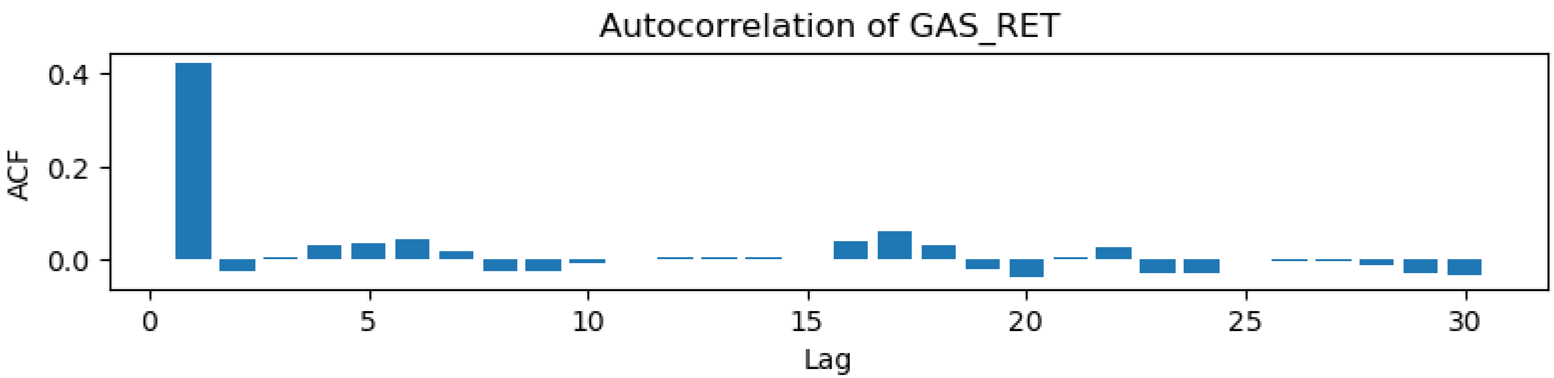

DE_COAL: time series with rolling means
DE_COAL
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_COAL
roll7
roll14
roll30

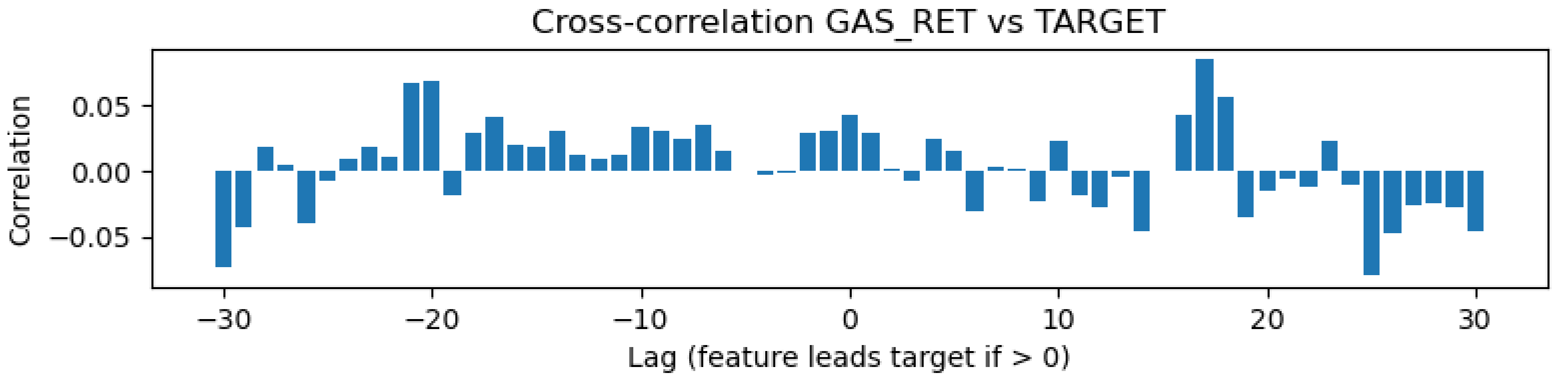

Autocorrelation of DE_COAL
ACF
0.2
0.0
0
5
10
15
20
25
30
Lag

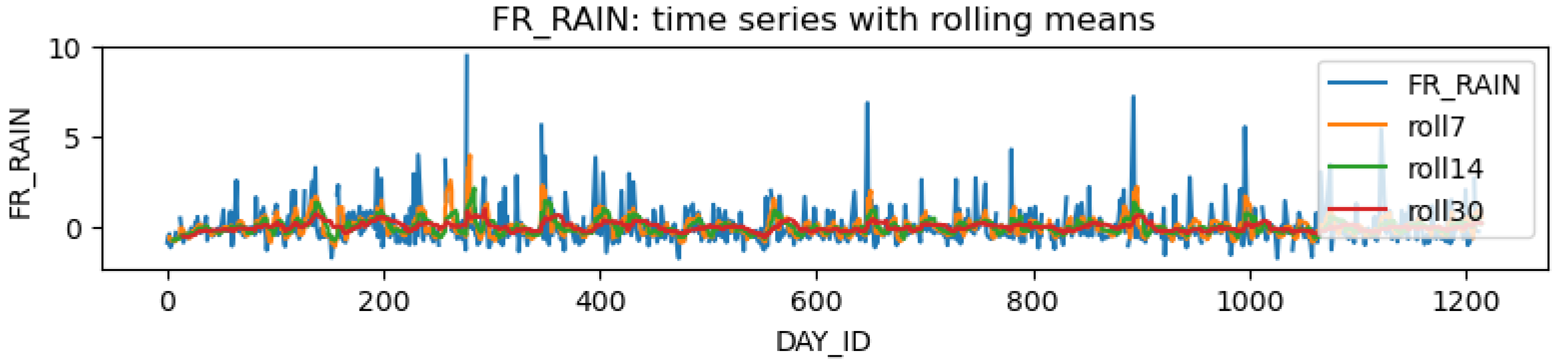

Cross-correlation DE_COAL vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)
DE_TEMP: time series with rolling means
DE_TEMP
2.5
0.0
−2.5
0
200
400
600
800
1000
1200
DAY_ID
DE_TEMP
roll7
roll14
roll30
Autocorrelation of DE_TEMP
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

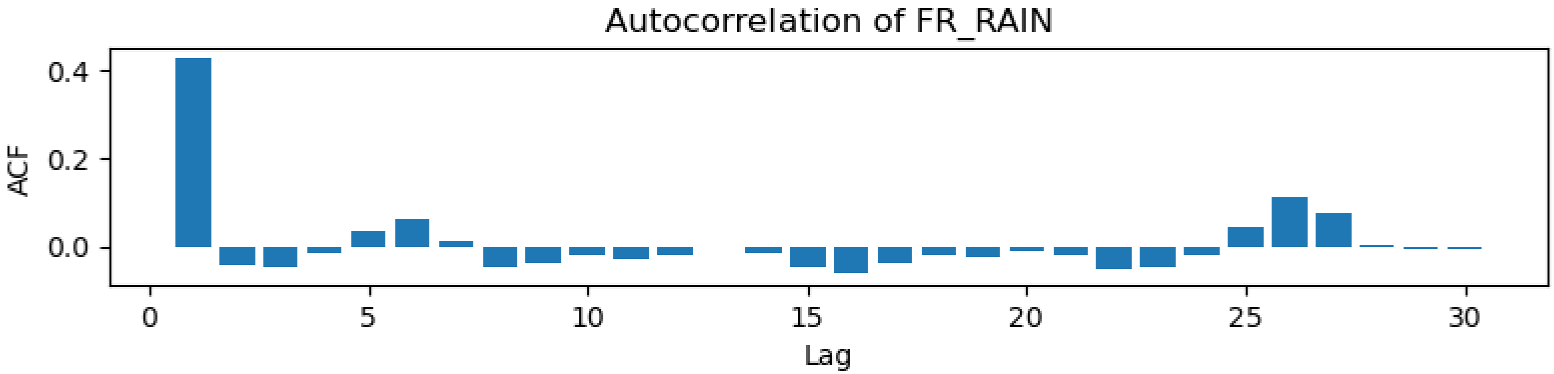

Cross-correlation DE_TEMP vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

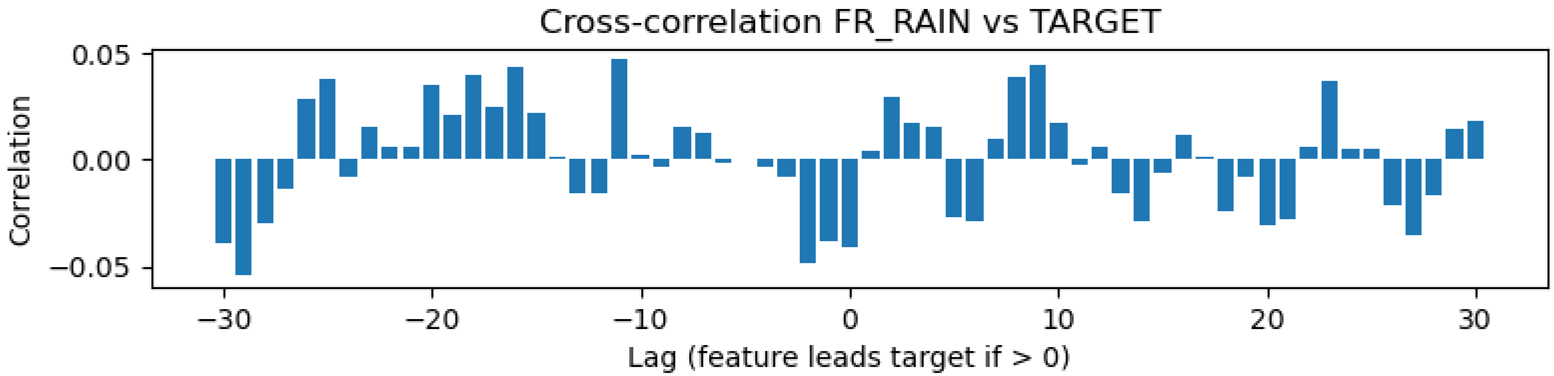

GAS_RET: time series with rolling means
GAS_RET
5
0
−5
0
200
400
600
800
1000
1200
DAY_ID
GAS_RET
roll7
roll14
roll30

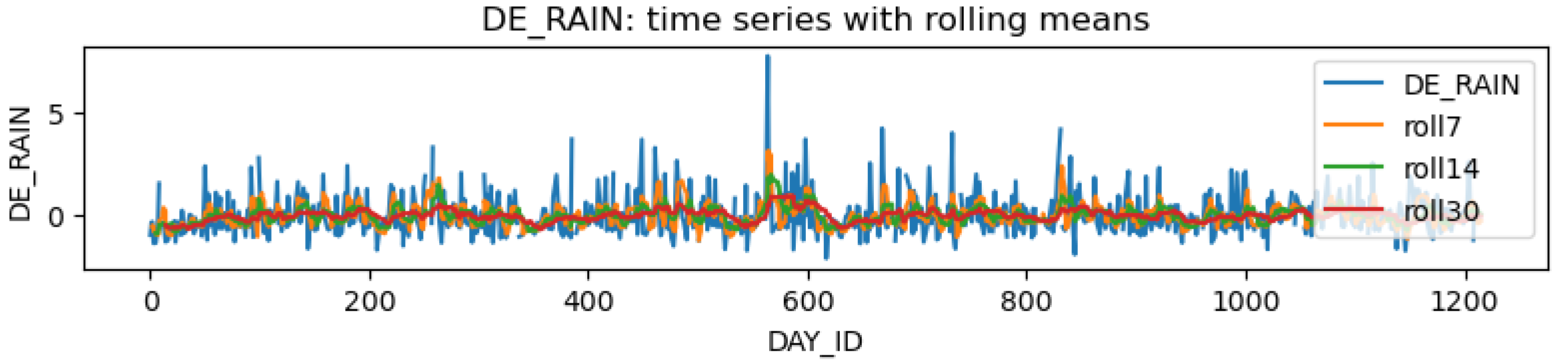

Autocorrelation of GAS_RET
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

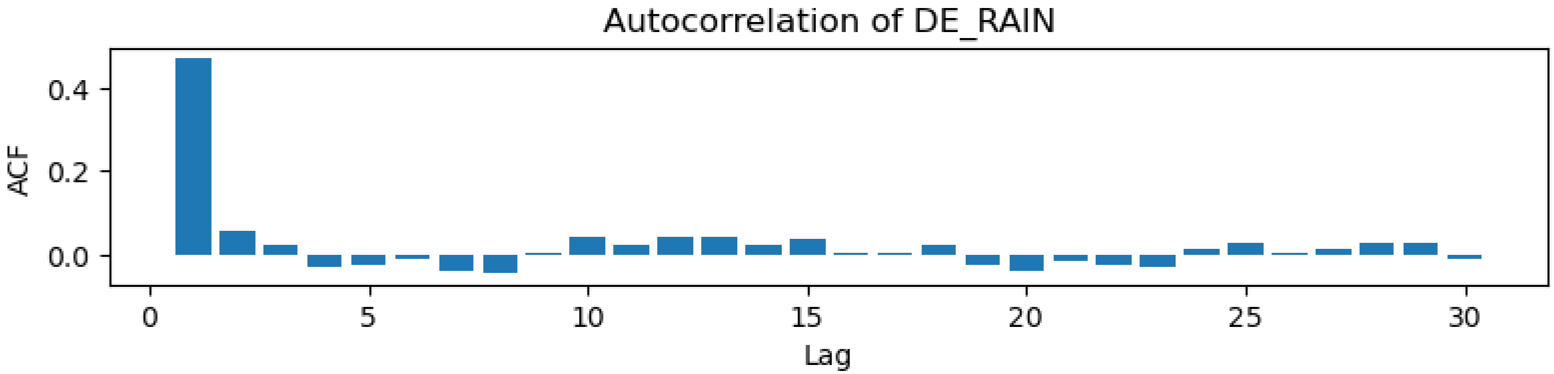

Cross-correlation GAS_RET vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

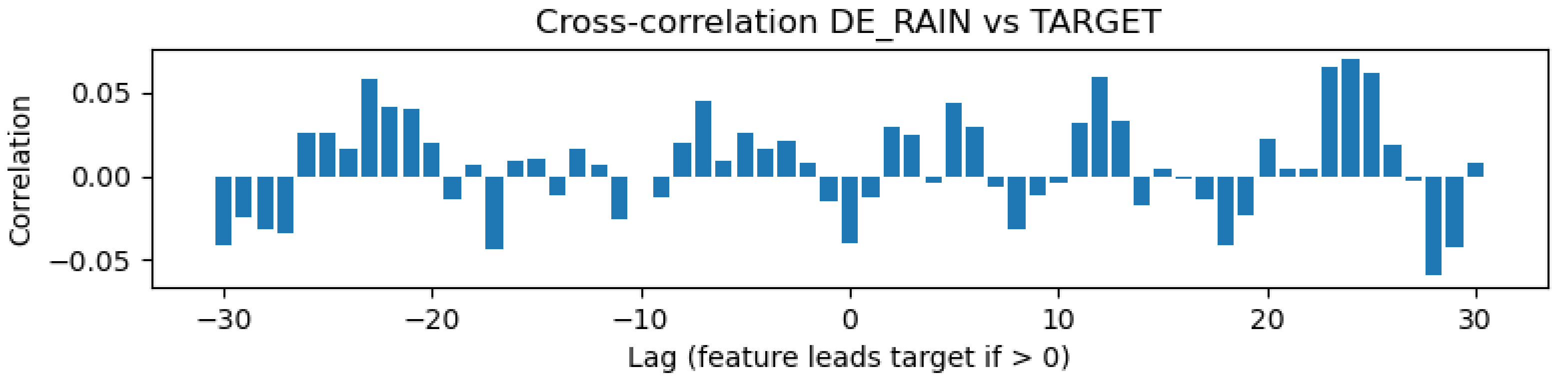

FR_RAIN: time series with rolling means
FR_RAIN
10
5
0
0
200
400
600
800
1000
1200
DAY_ID
FR_RAIN
roll7
roll14
roll30

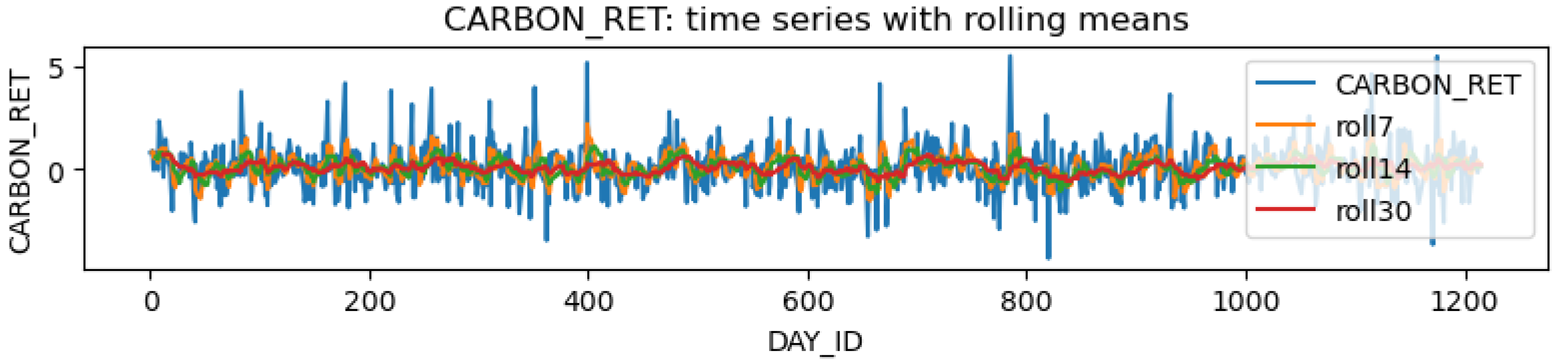

Autocorrelation of FR_RAIN
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

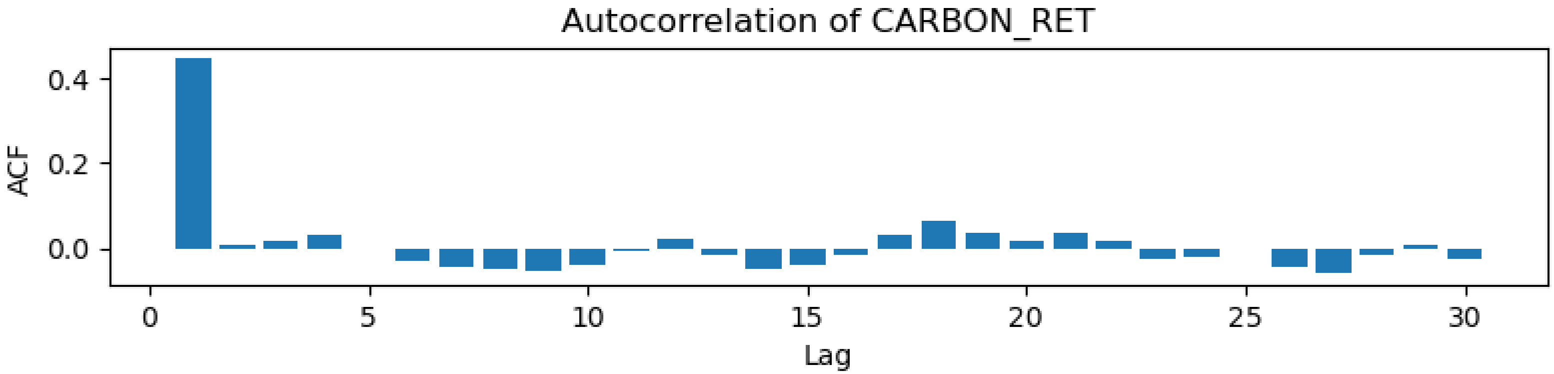

Cross-correlation FR_RAIN vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

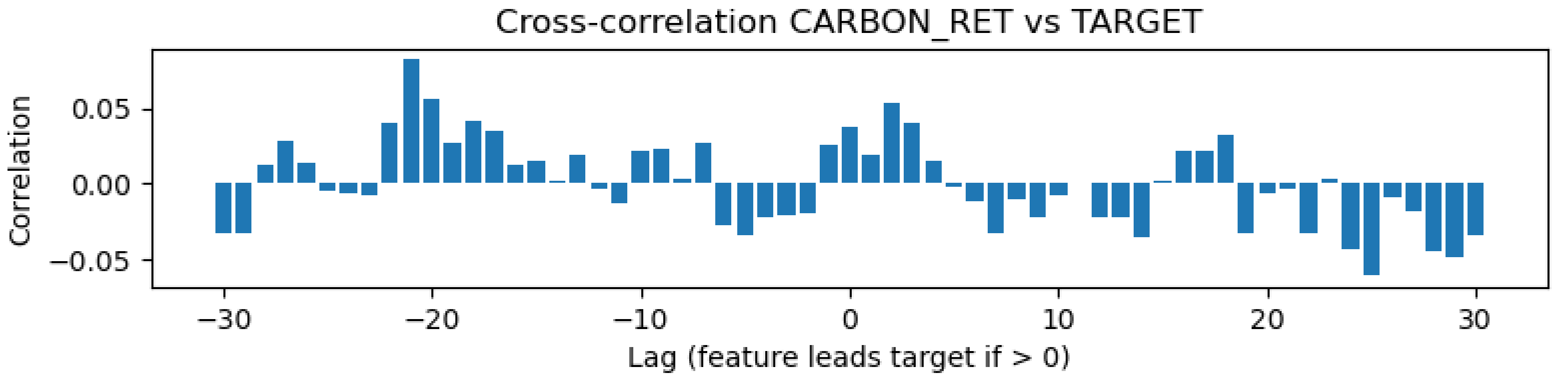

DE_RAIN: time series with rolling means
DE_RAIN
5
0
0
200
400
600
800
1000
1200
DAY_ID
DE_RAIN
roll7
roll14
roll30

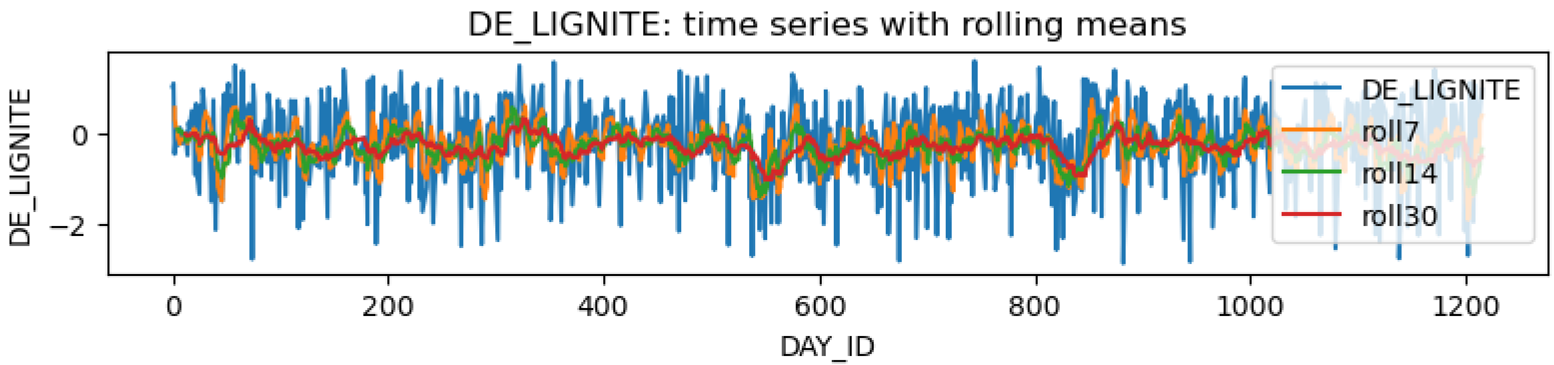

Autocorrelation of DE_RAIN
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

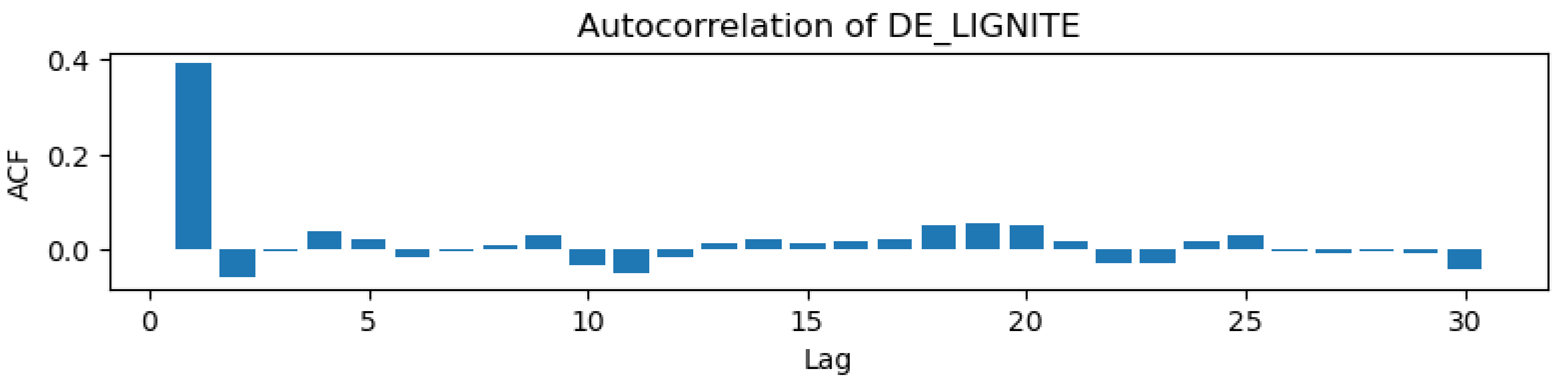

Cross-correlation DE_RAIN vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

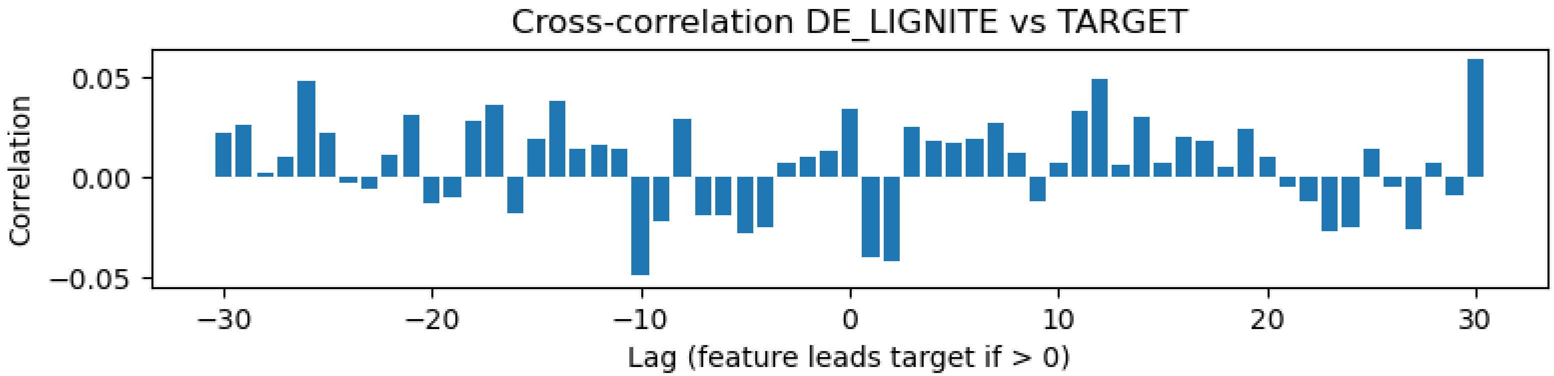

CARBON_RET: time series with rolling means
CARBON_RET
5
0
0
200
400
600
800
1000
1200
DAY_ID
CARBON_RET
roll7
roll14
roll30

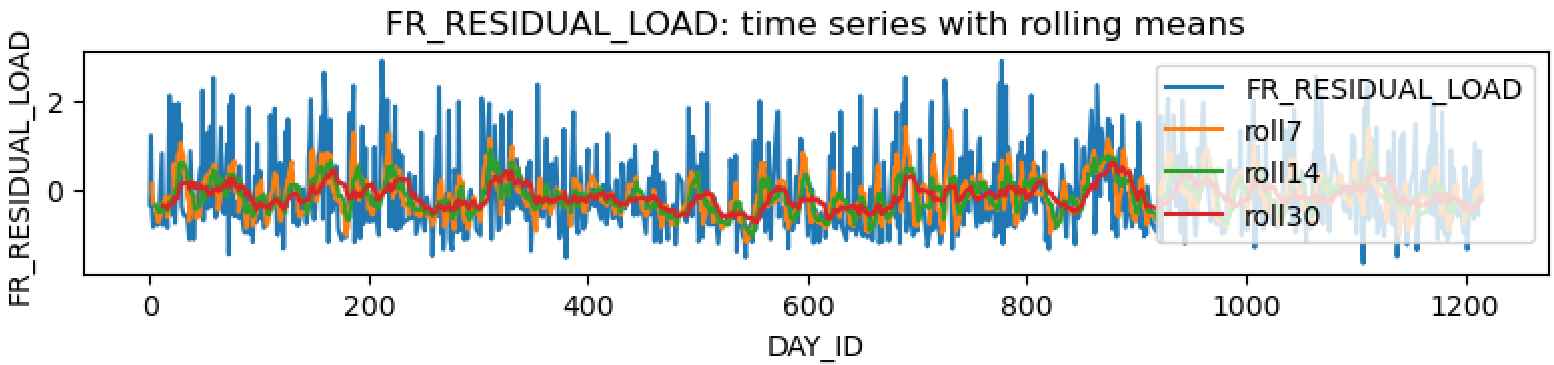

Autocorrelation of CARBON_RET
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

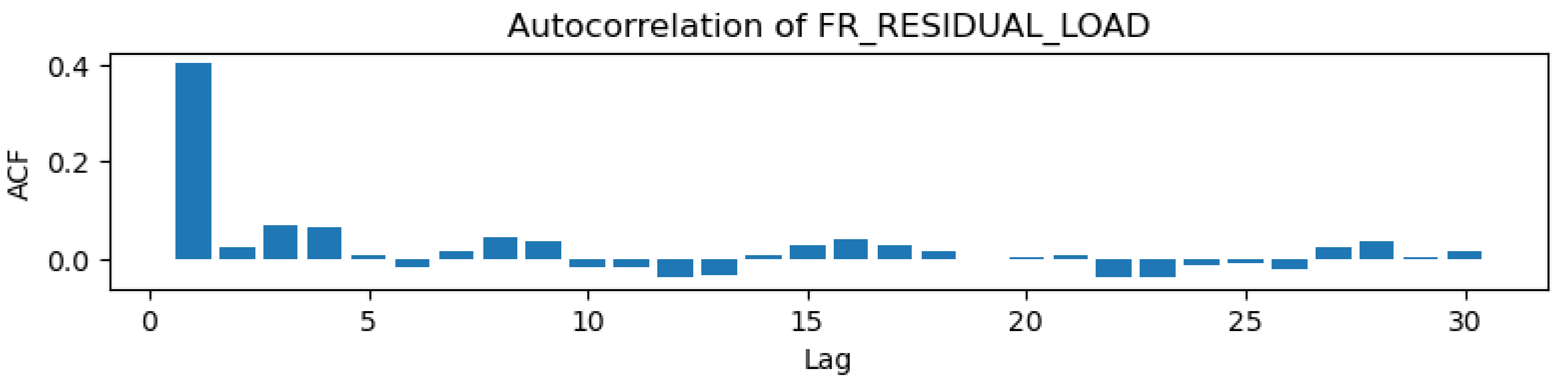

Cross-correlation CARBON_RET vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

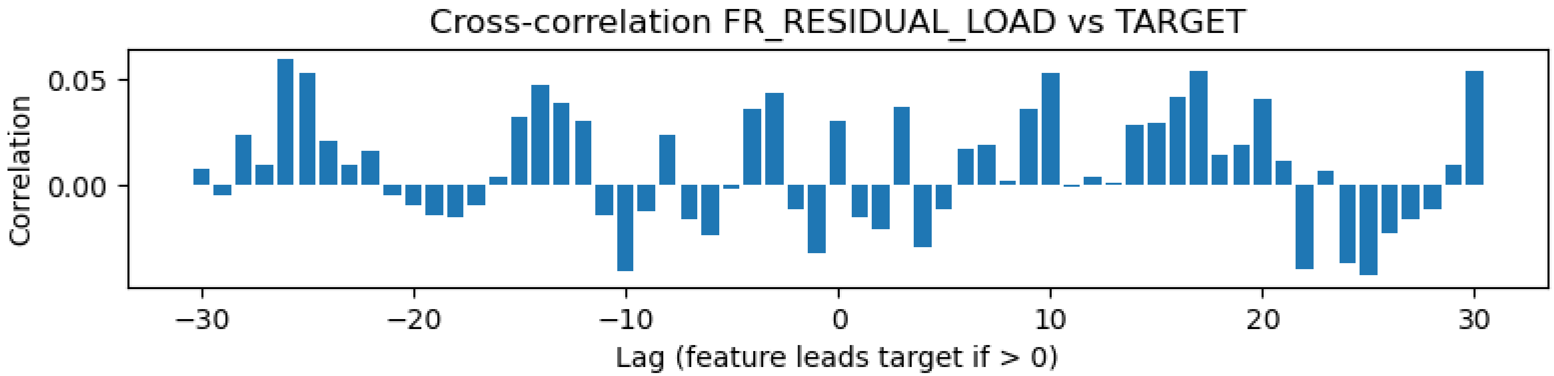

DE_LIGNITE: time series with rolling means
DE_LIGNITE
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_LIGNITE
roll7
roll14
roll30

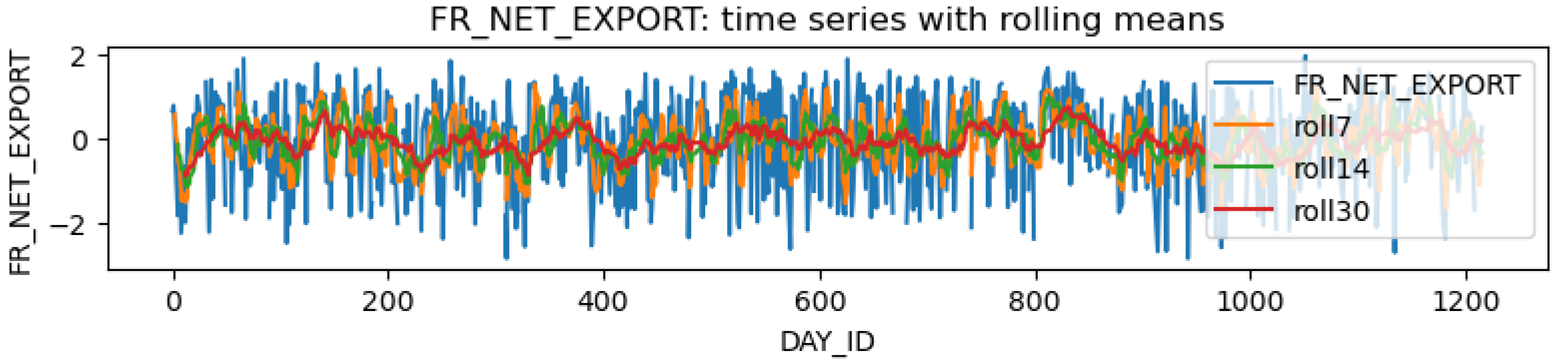

Autocorrelation of DE_LIGNITE
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

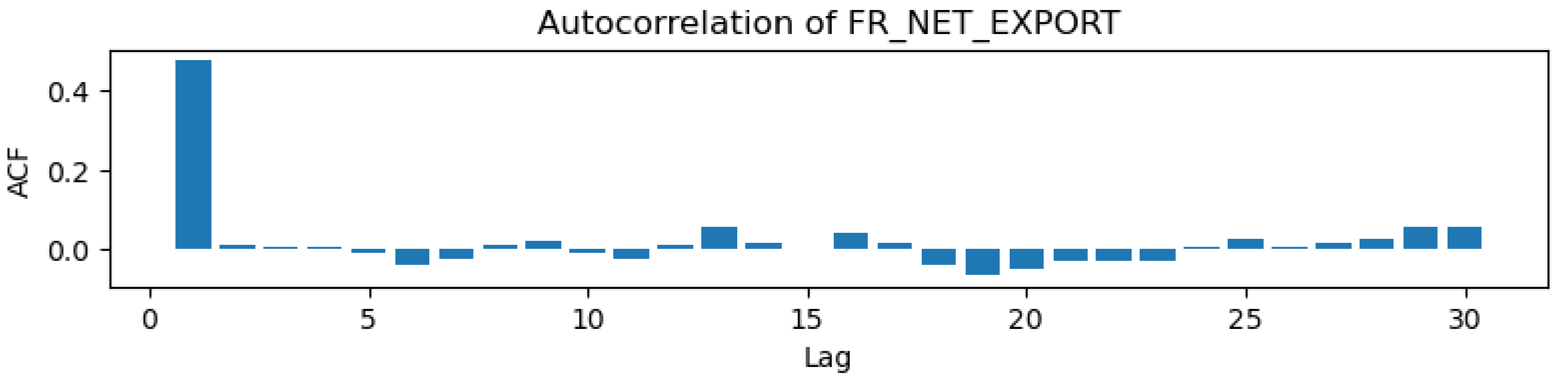

Cross-correlation DE_LIGNITE vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

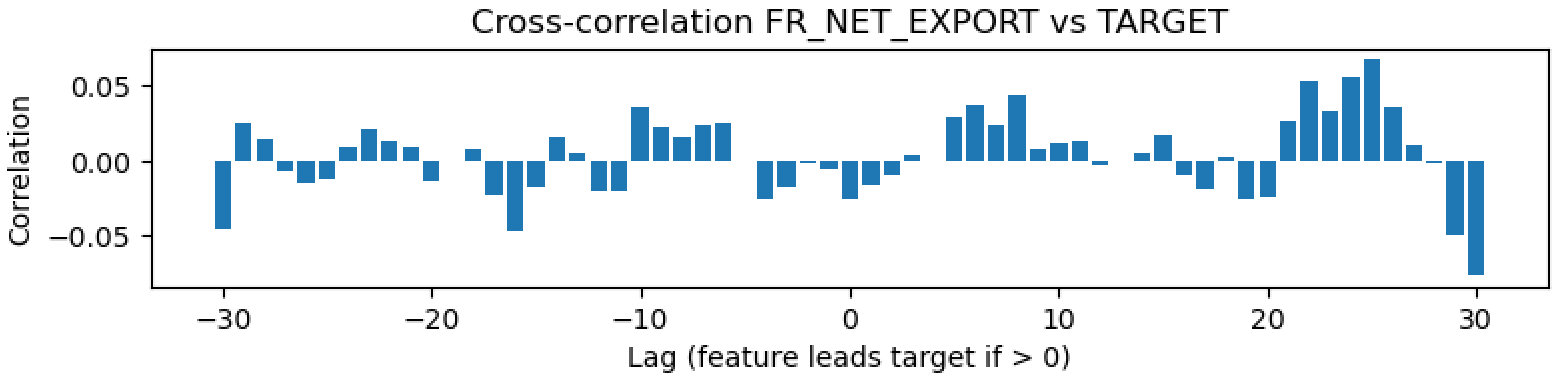

FR_RESIDUAL_LOAD: time series with rolling means
FR_RESIDUAL_LOAD
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_RESIDUAL_LOAD
roll7
roll14
roll30

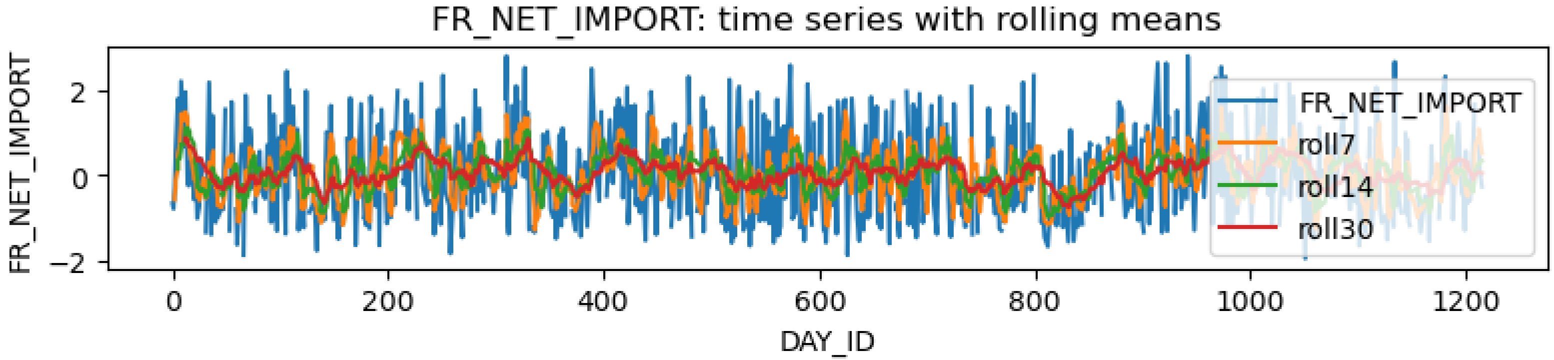

Autocorrelation of FR_RESIDUAL_LOAD
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

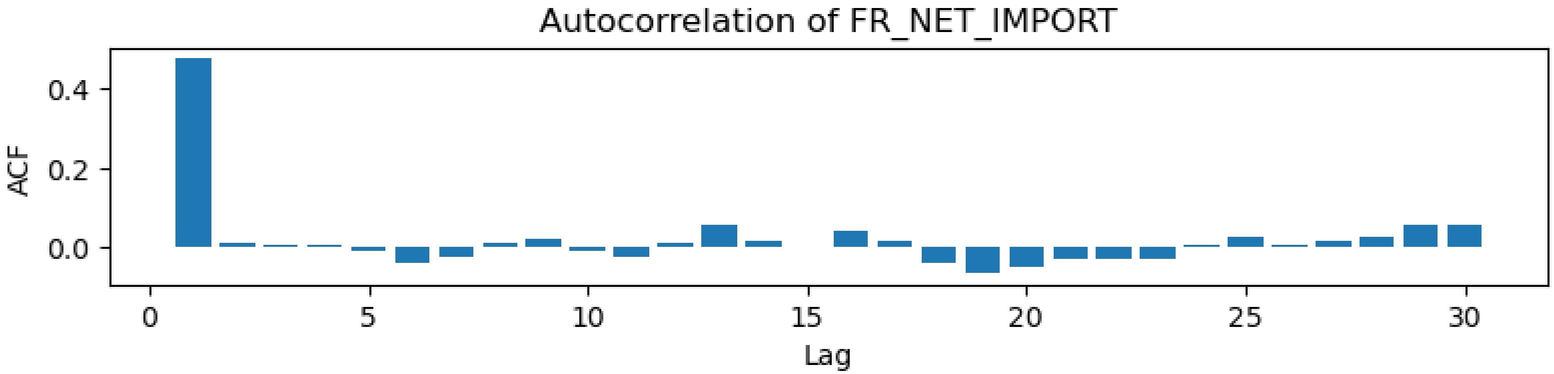

Cross-correlation FR_RESIDUAL_LOAD vs TARGET
Correlation
0.05
0.00
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

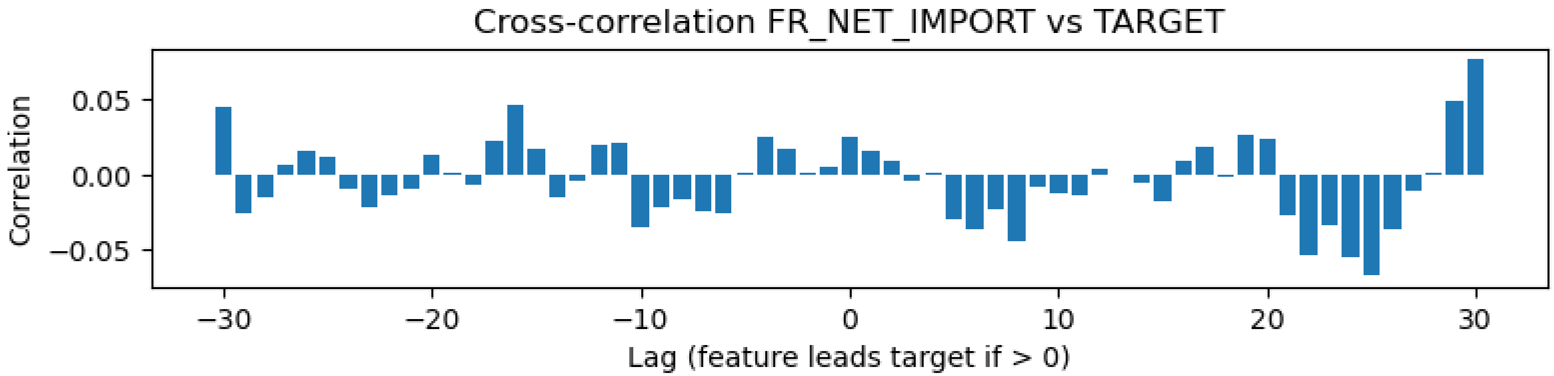

FR_NET_EXPORT: time series with rolling means
FR_NET_EXPORT
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
FR_NET_EXPORT
roll7
roll14
roll30

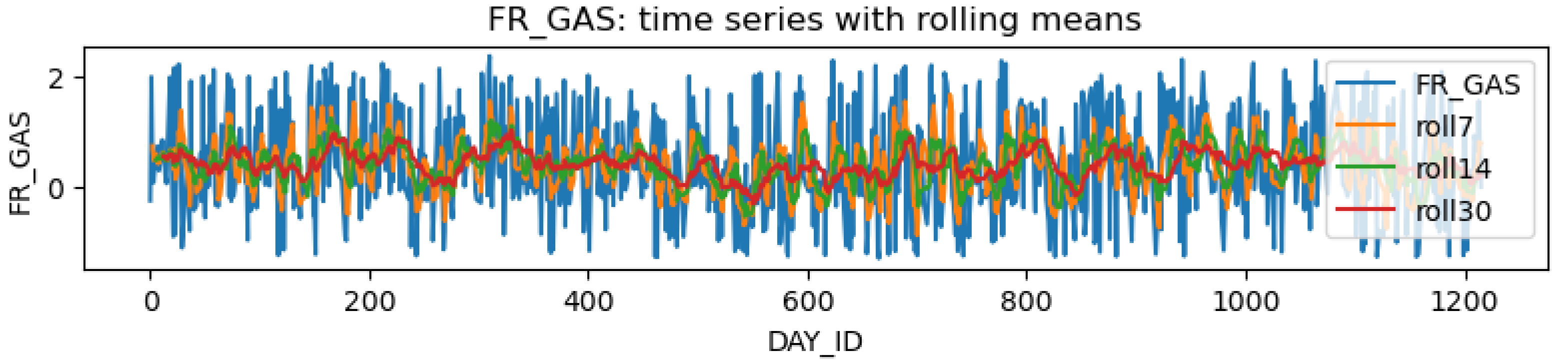

Autocorrelation of FR_NET_EXPORT
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

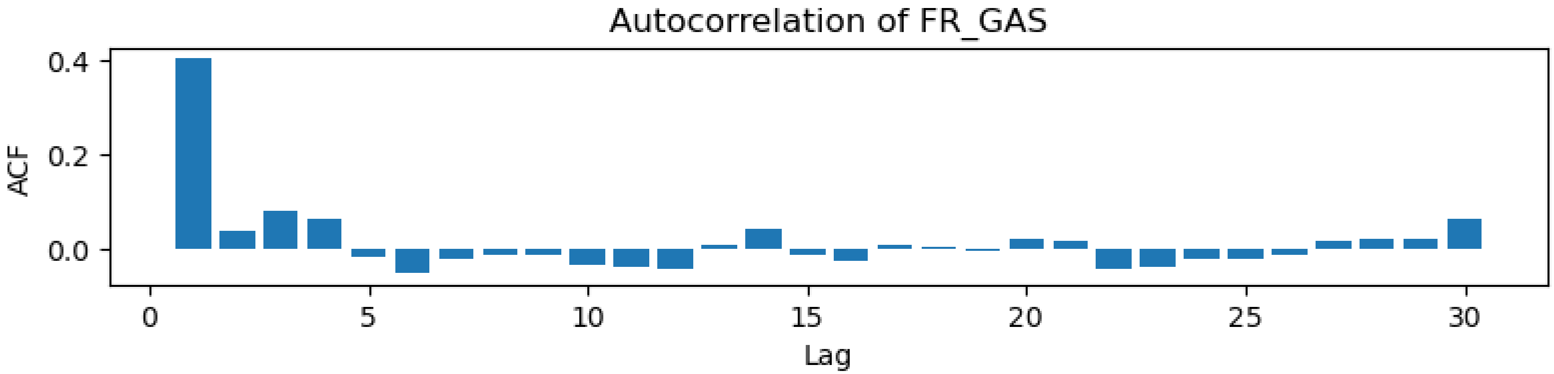

Cross-correlation FR_NET_EXPORT vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

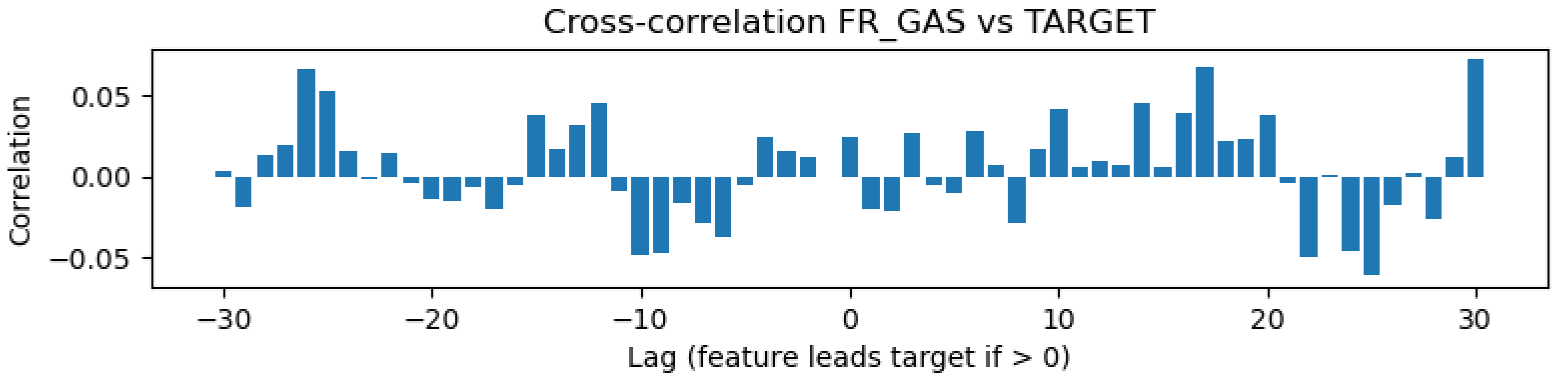

FR_NET_IMPORT: time series with rolling means
FR_NET_IMPORT
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
FR_NET_IMPORT
roll7
roll14
roll30

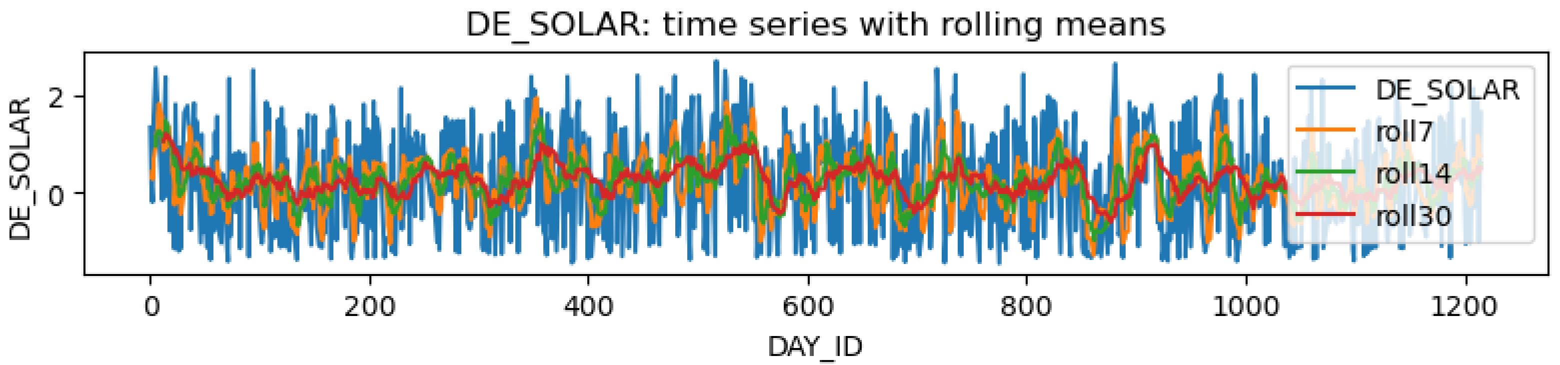

Autocorrelation of FR_NET_IMPORT
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

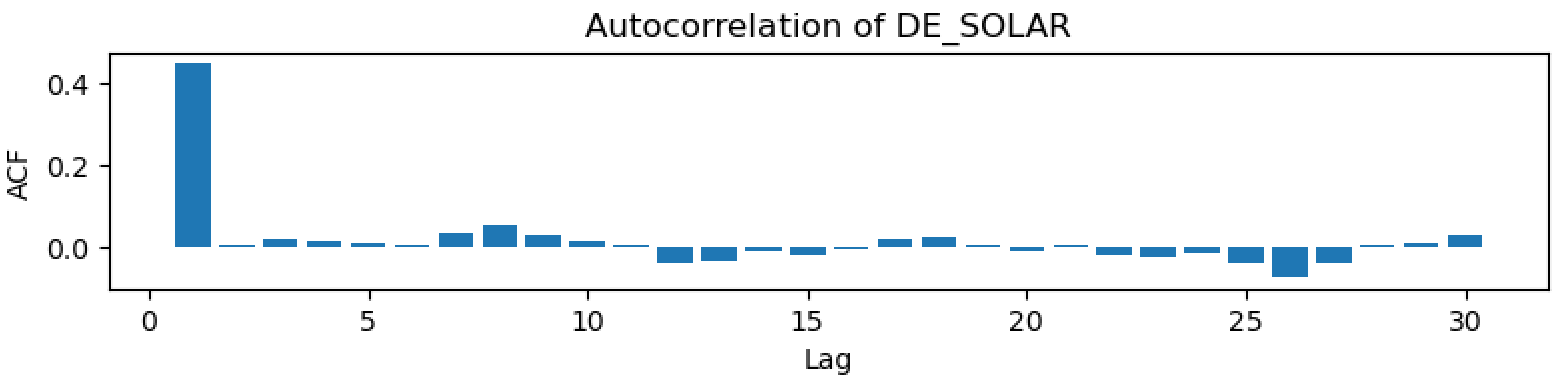

Cross-correlation FR_NET_IMPORT vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

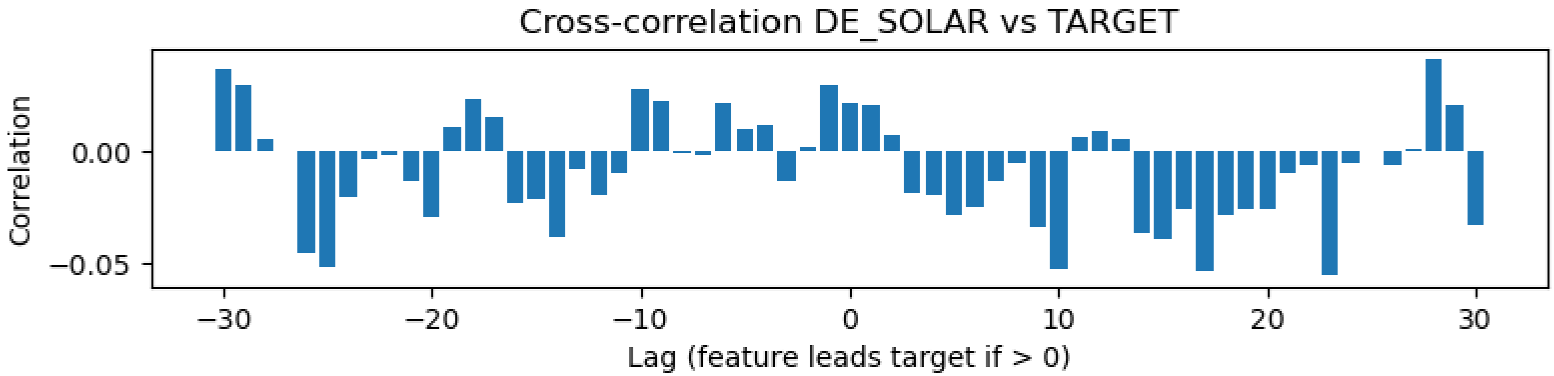

FR_GAS: time series with rolling means
FR_GAS
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_GAS
roll7
roll14
roll30

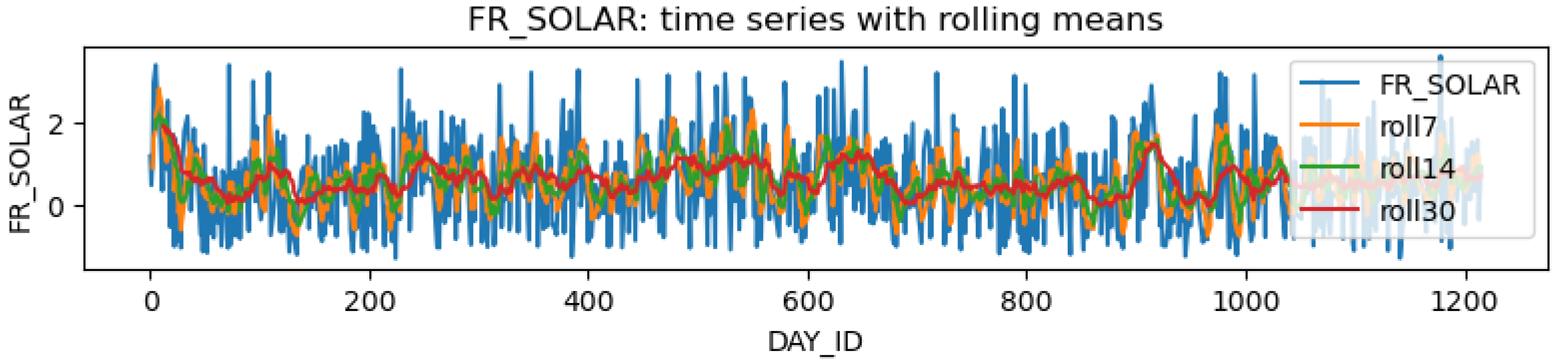

Autocorrelation of FR_GAS
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

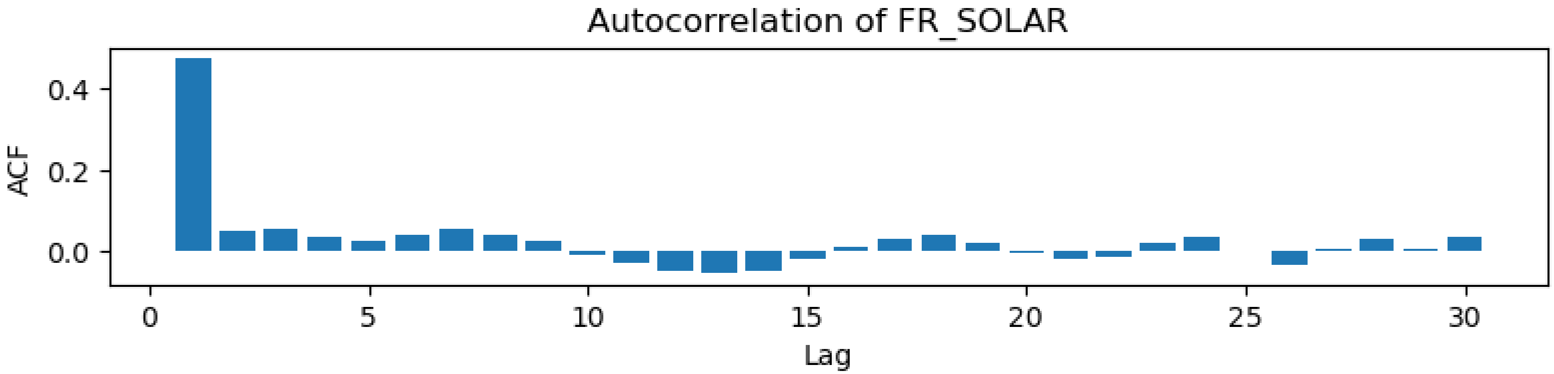

Cross-correlation FR_GAS vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

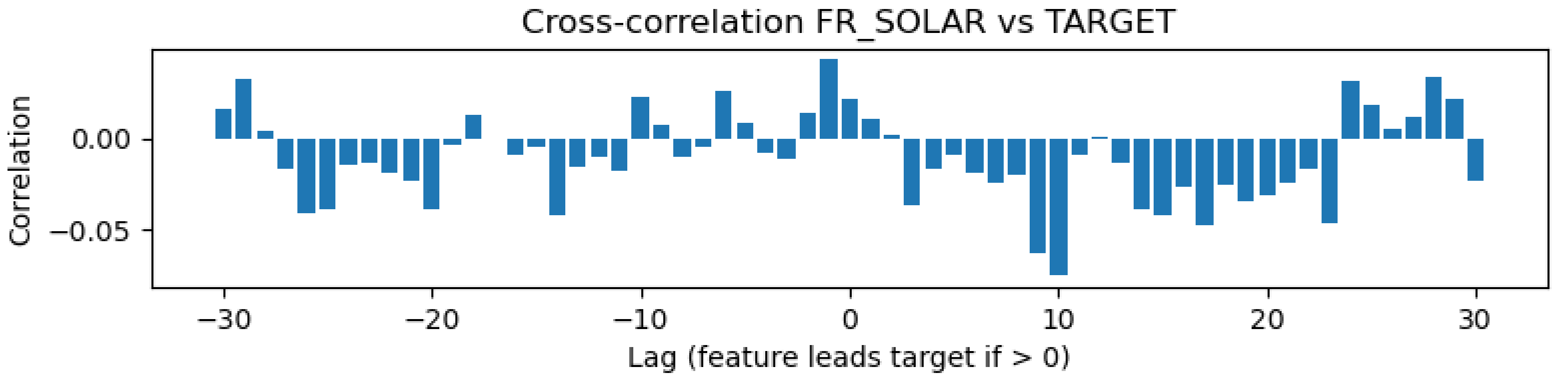

DE_SOLAR: time series with rolling means
DE_SOLAR
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_SOLAR
roll7
roll14
roll30

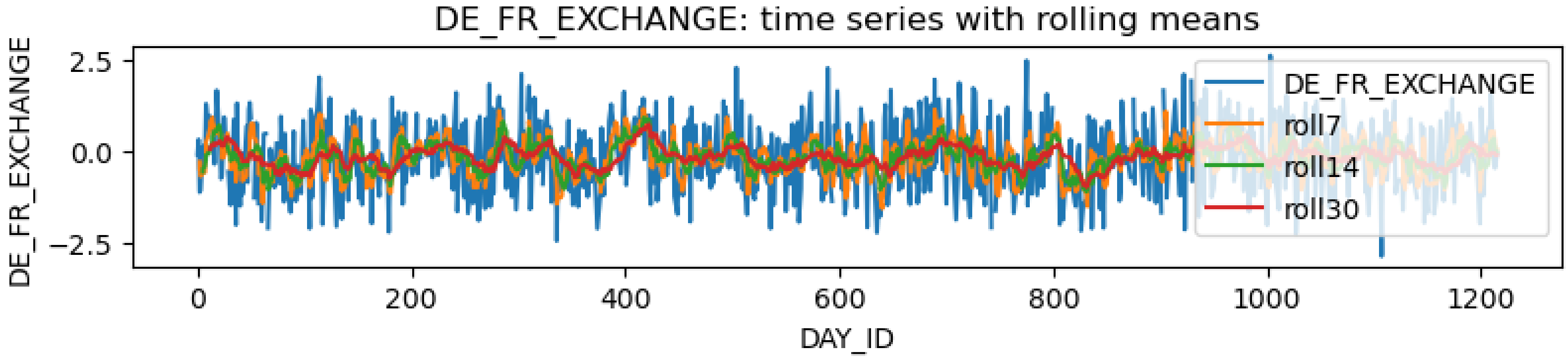

Autocorrelation of DE_SOLAR
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

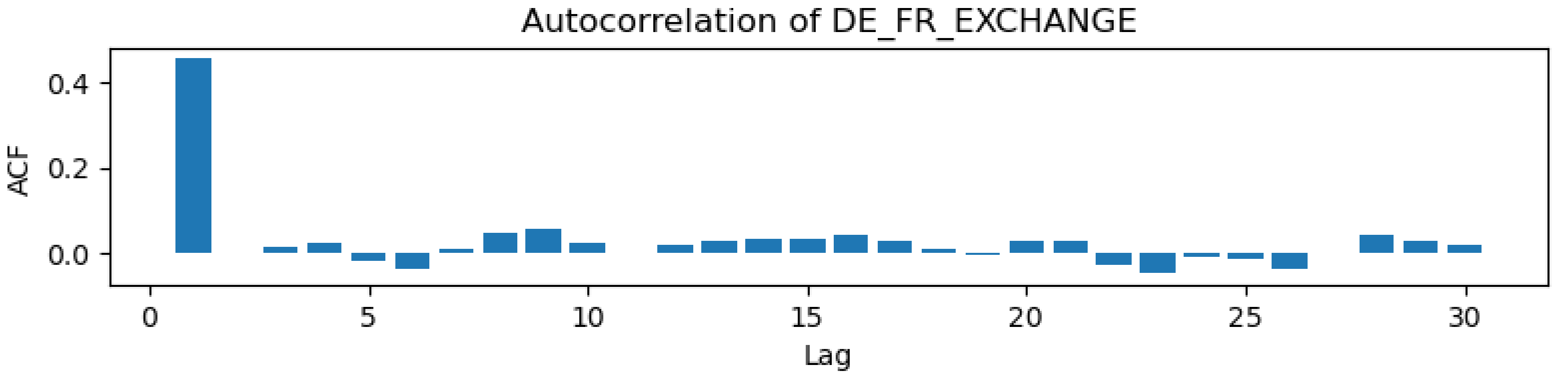

Cross-correlation DE_SOLAR vs TARGET
Correlation
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

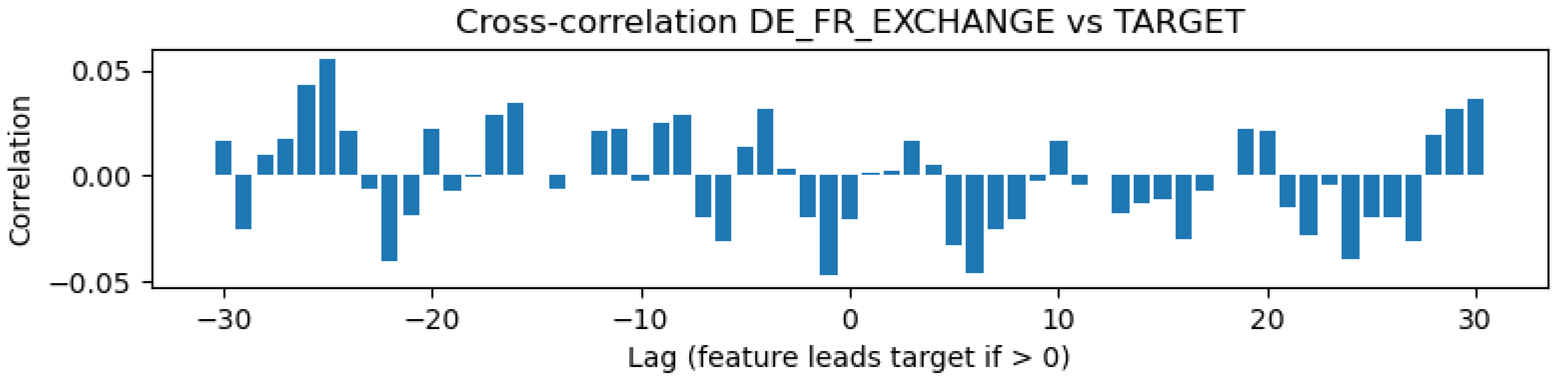

FR_SOLAR: time series with rolling means
FR_SOLAR
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_SOLAR
roll7
roll14
roll30

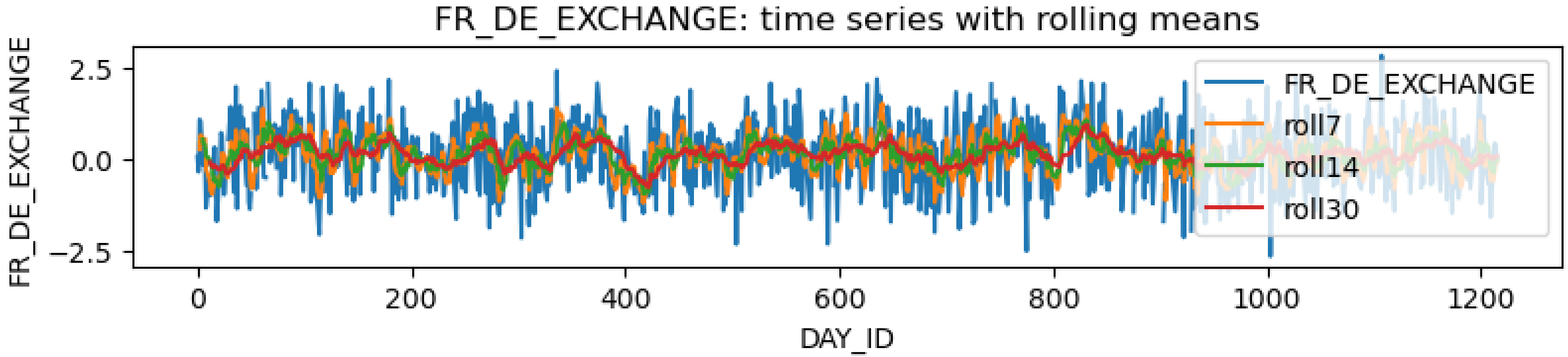

Autocorrelation of FR_SOLAR
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

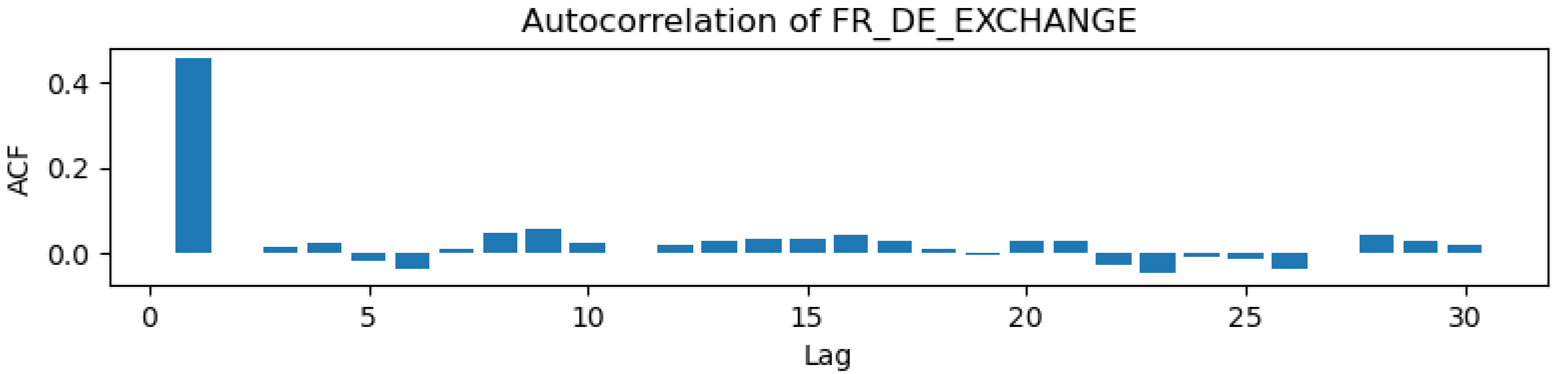

Cross-correlation FR_SOLAR vs TARGET
Correlation
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

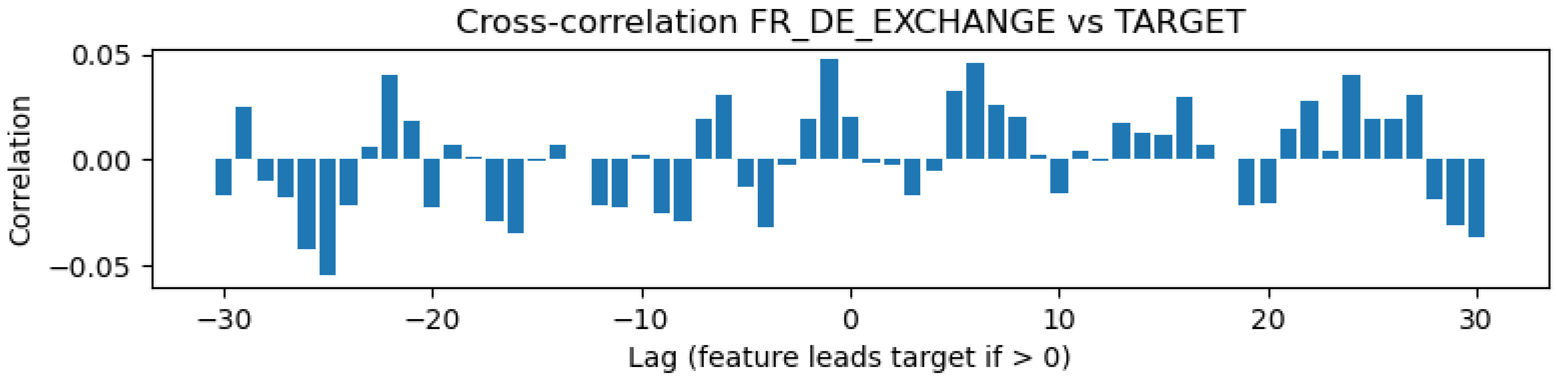

DE_FR_EXCHANGE: time series with rolling means
DE_FR_EXCHANGE
2.5
0.0
−2.5
0
200
400
600
800
1000
1200
DAY_ID
DE_FR_EXCHANGE
roll7
roll14
roll30

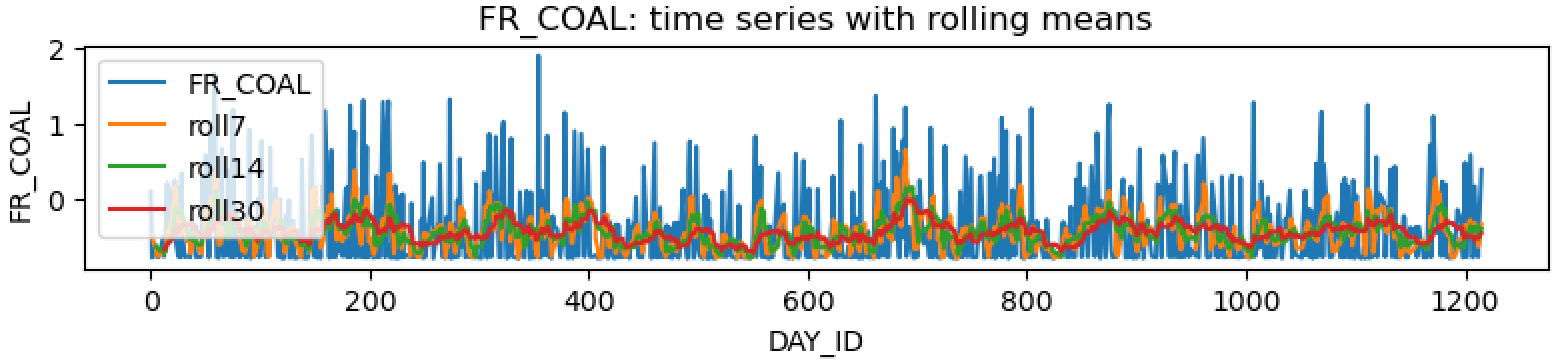

Autocorrelation of DE_FR_EXCHANGE
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

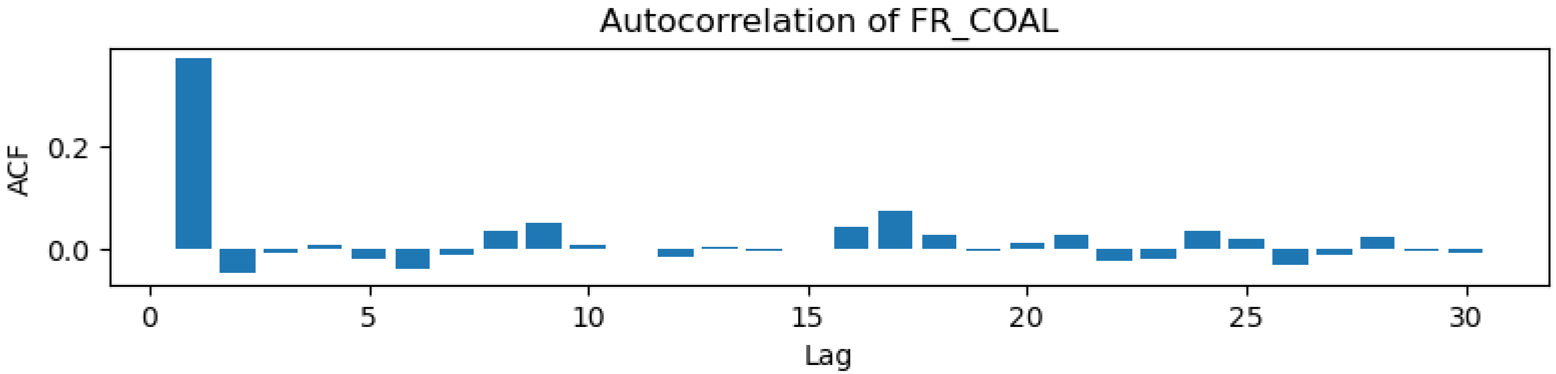

Cross-correlation DE_FR_EXCHANGE vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

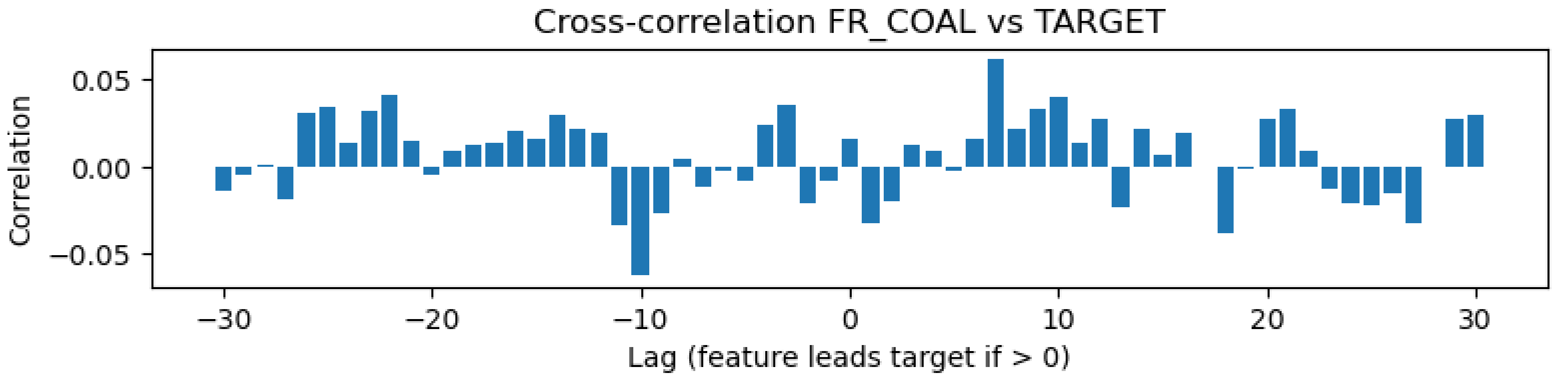

FR_DE_EXCHANGE: time series with rolling means
FR_DE_EXCHANGE
2.5
0.0
−2.5
0
200
400
600
800
1000
1200
DAY_ID
FR_DE_EXCHANGE
roll7
roll14
roll30

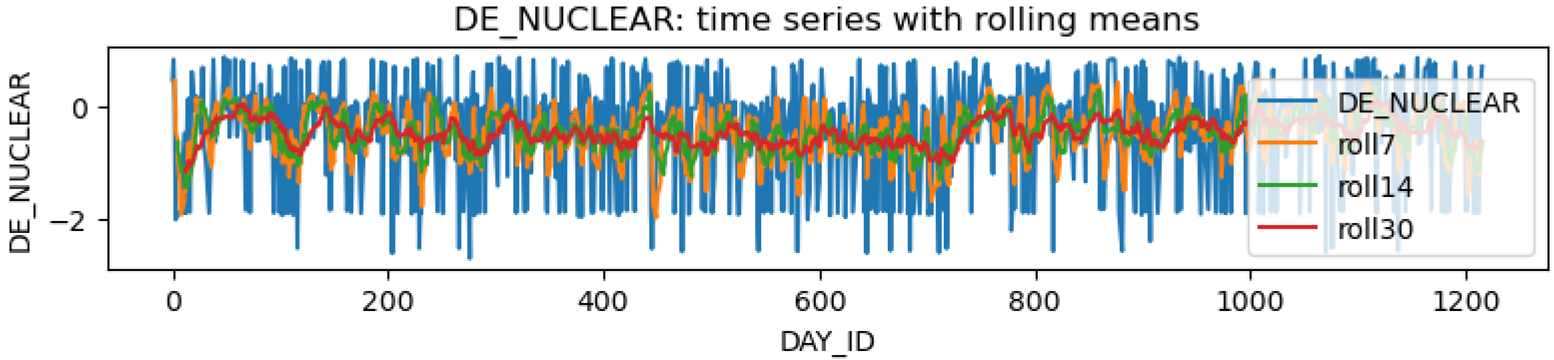

Autocorrelation of FR_DE_EXCHANGE
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

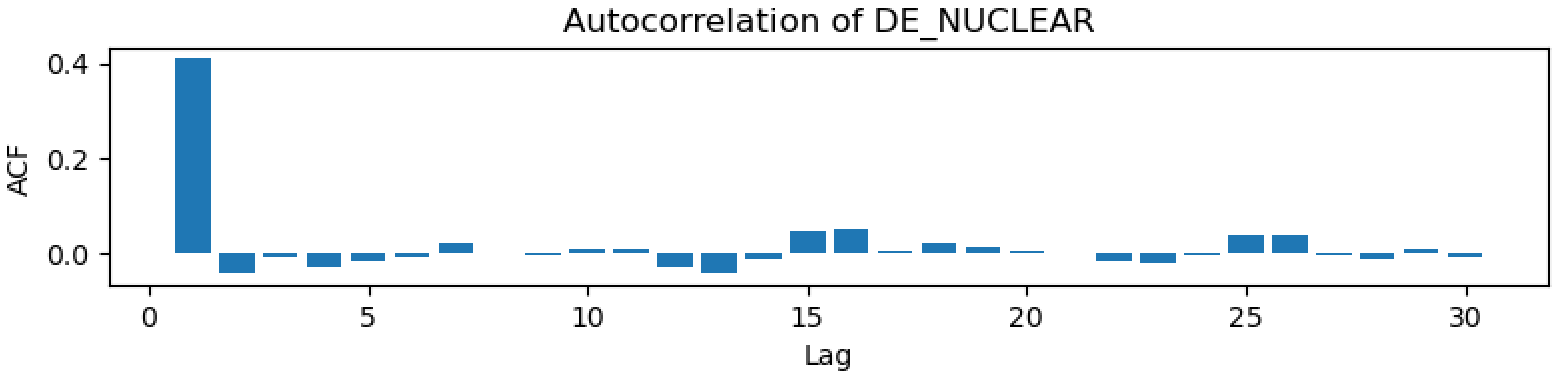

Cross-correlation FR_DE_EXCHANGE vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

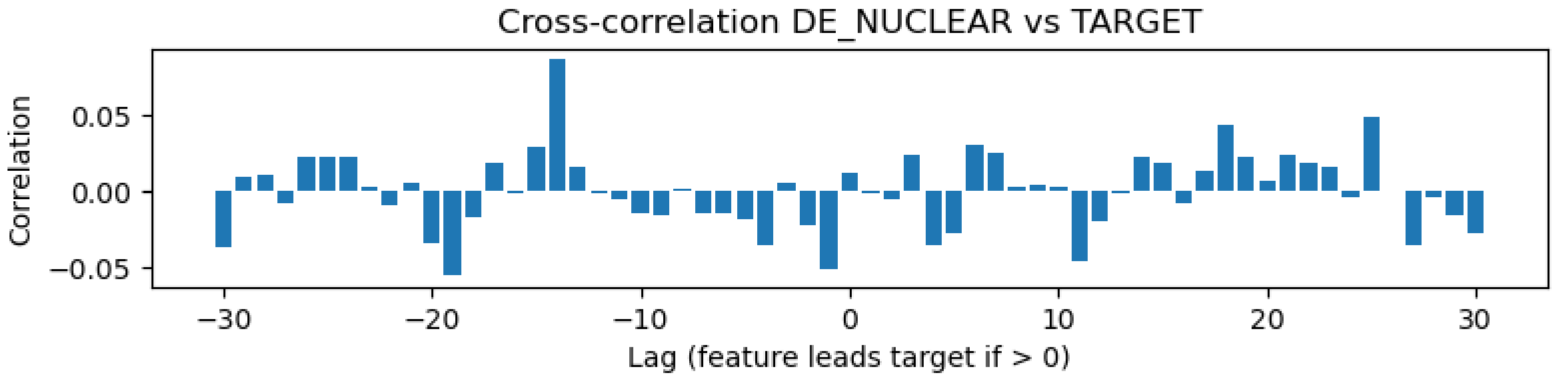

FR_COAL: time series with rolling means
FR_COAL
roll7
roll14
roll30
2
1
0
0
200
400
600
800
1000
1200
DAY_ID

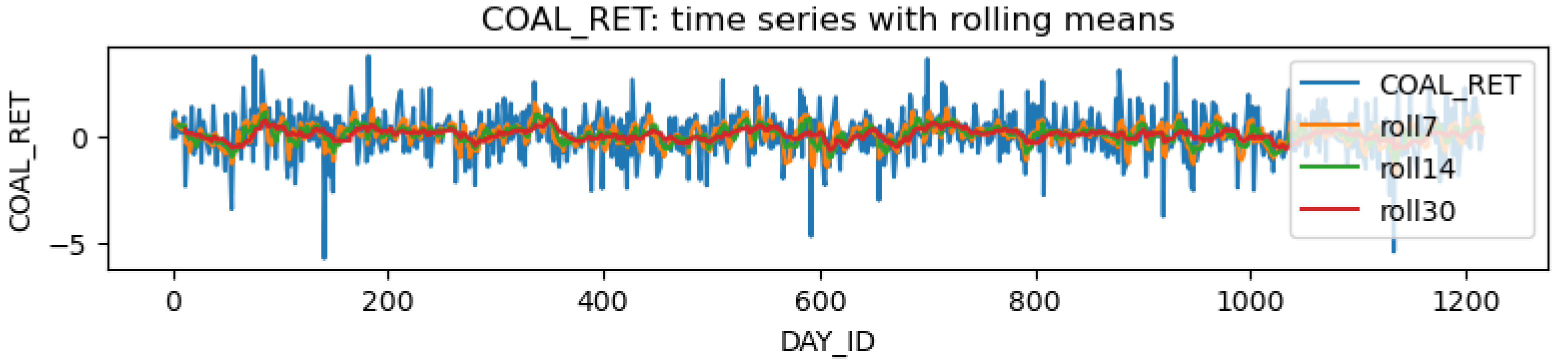

Autocorrelation of FR_COAL
ACF
0.2
0.0
0
5
10
15
20
25
30
Lag

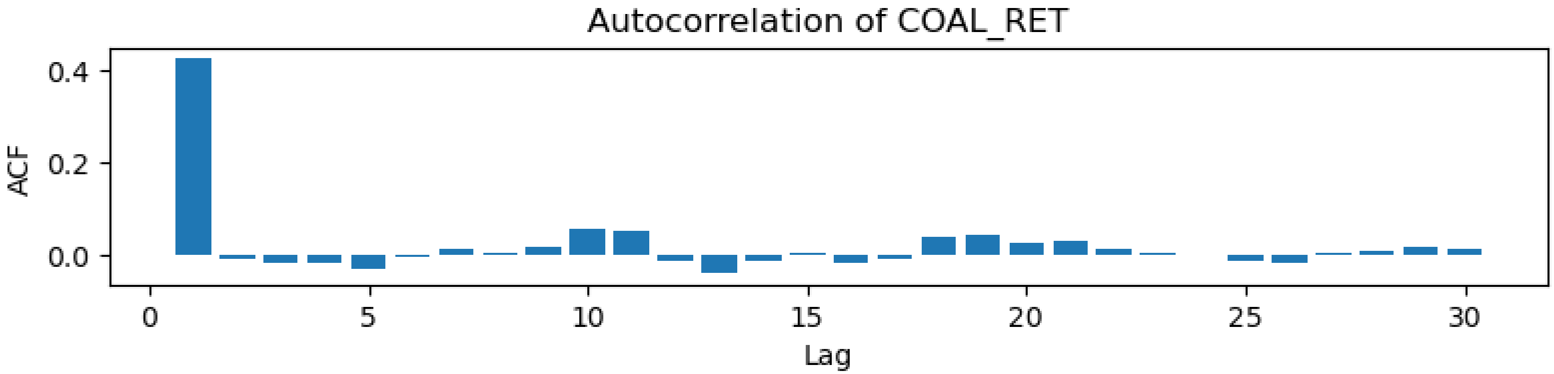

Cross-correlation FR_COAL vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

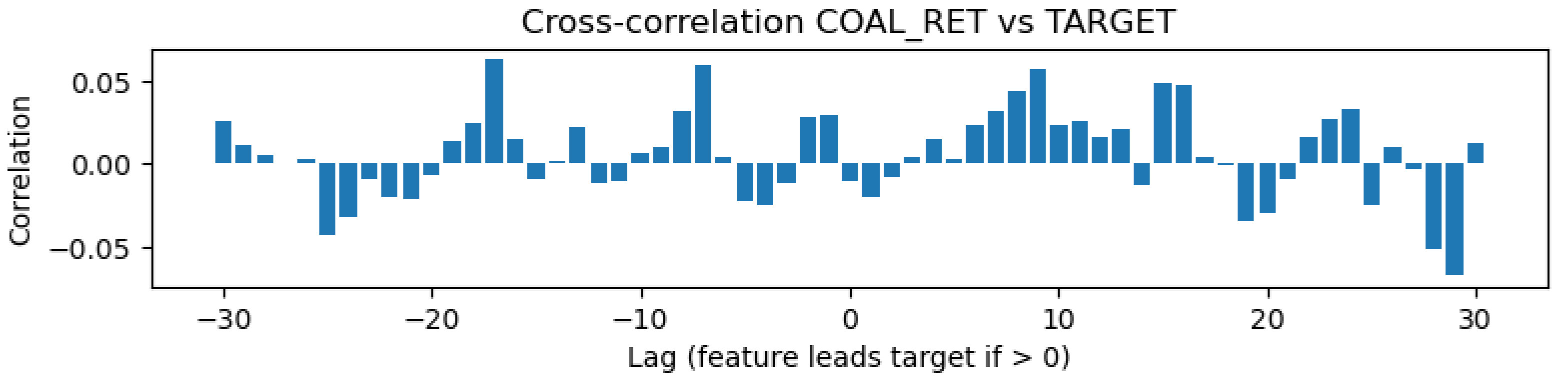

DE_NUCLEAR: time series with rolling means
DE_NUCLEAR
0
−2
DE_NUCLEAR
roll7
roll14
roll30
0
200
400
600
800
1000
1200
DAY_ID

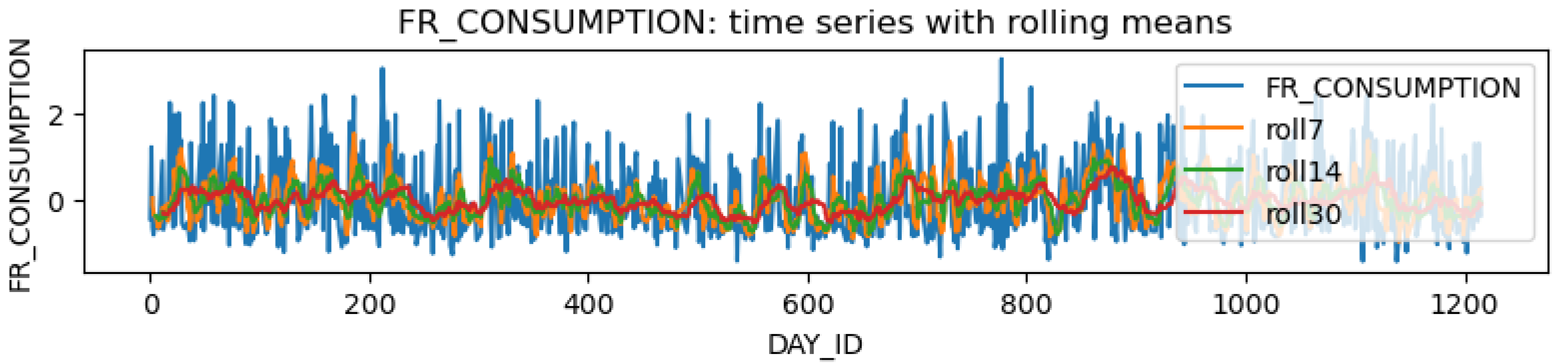

Autocorrelation of DE_NUCLEAR
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

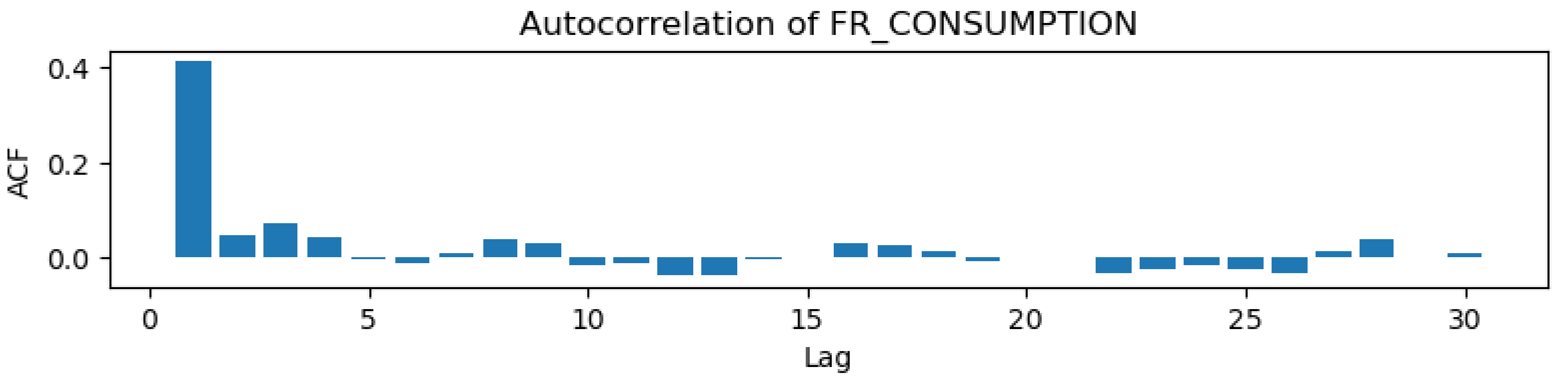

Cross-correlation DE_NUCLEAR vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

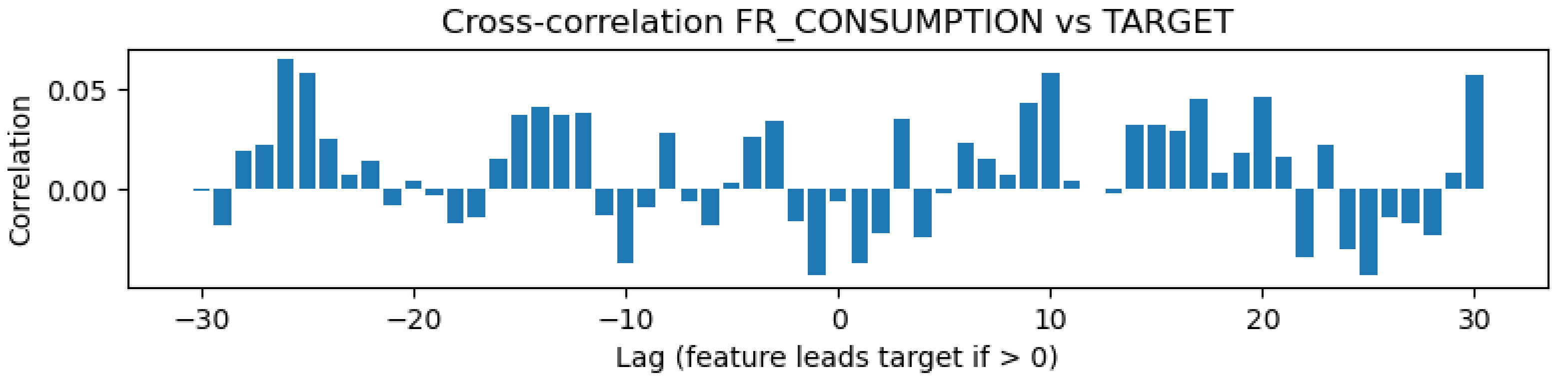

COAL_RET: time series with rolling means
COAL_RET
0
−5
0
200
400
600
800
1000
1200
DAY_ID
COAL_RET
roll7
roll14
roll30

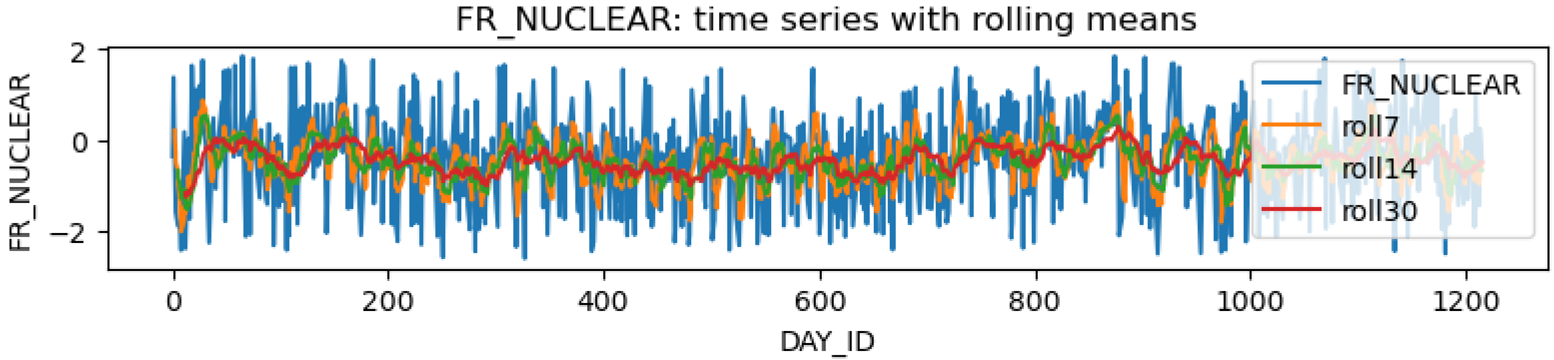

Autocorrelation of COAL_RET
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

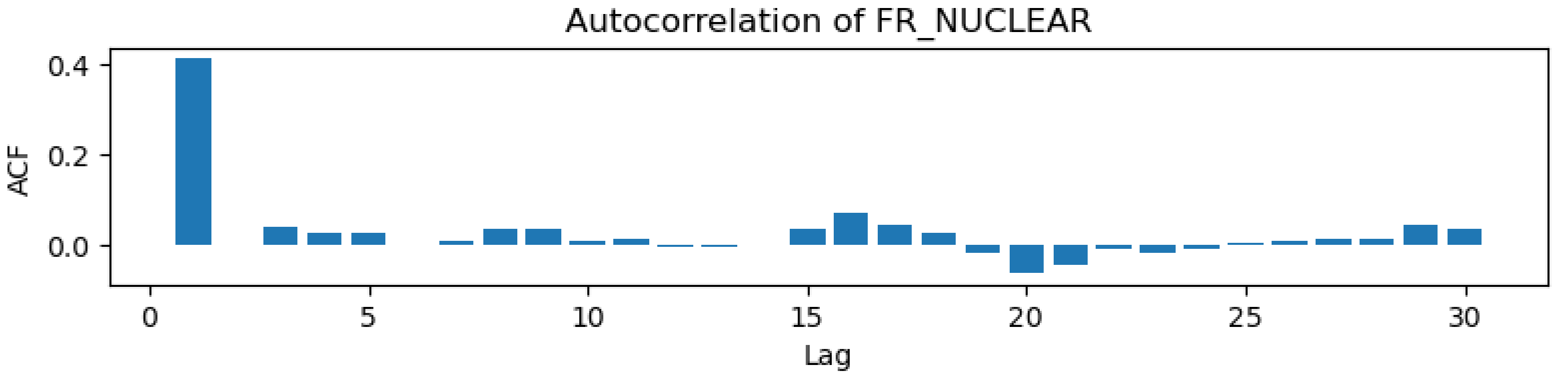

Cross-correlation COAL_RET vs TARGET
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

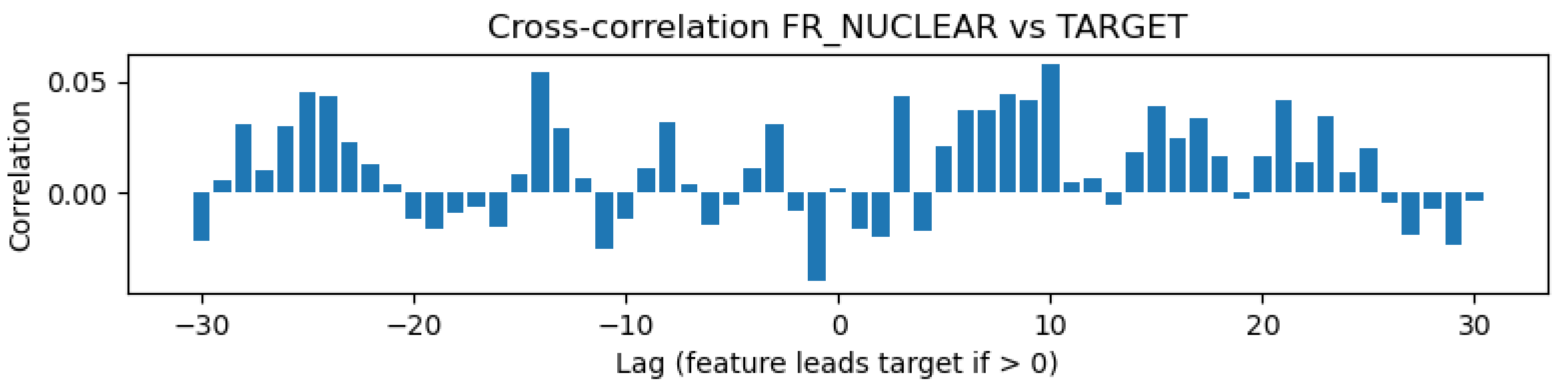

FR_CONSUMPTION: time series with rolling means
FR_CONSUMPTION
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_CONSUMPTION
roll7
roll14
roll30


Autocorrelation of FR_CONSUMPTION
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag


Cross-correlation FR_CONSUMPTION vs TARGET
Correlation
0.05
0.00
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)


FR_NUCLEAR: time series with rolling means
FR_NUCLEAR
2
0
−2
FR_NUCLEAR
roll7
roll14
roll30
0
200
400
600
800
1000
1200
DAY_ID


Autocorrelation of FR_NUCLEAR
ACF
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag



8. Per-Country EDA (FR / DE)

In [17]:
```
if "COUNTRY" not in df.columns:
    print("COUNTRY column not found; skipping per-country EDA.")
else:
    countries = [c for c in df["COUNTRY"].dropna().unique()]
    print("Countries found:", countries)

    for cc in countries:
        print("\n==============================")
        print(f"### COUNTRY: {cc}")
        d = df[df["COUNTRY"] == cc].copy()

        # Numeric features in this slice
        numc0 = [c for c in d.columns if c not in {"ID","DAY_ID","COUNTRY","TARGET"} and pd.api.types.is_numeric_dtype(d[c])]
        numc = [c for c in d.columns if c not in {"ID","COUNTRY","TARGET"} and pd.api.types.is_numeric_dtype(d[c])]
        print(f"Rows: {len(d)}; Numeric features: {len(numc)}")

        # Missingness (top 20)
        miss_c = d[numc].isna().sum().sort_values(ascending=False)
        miss_p = (miss_c / len(d)).sort_values(ascending=False)
        mdf = pd.DataFrame({"missing_count": miss_c, "missing_pct": miss_p})
        miss_count = (mdf['missing_count'] > 0).sum()
        display(mdf)

        plt.figure()
```

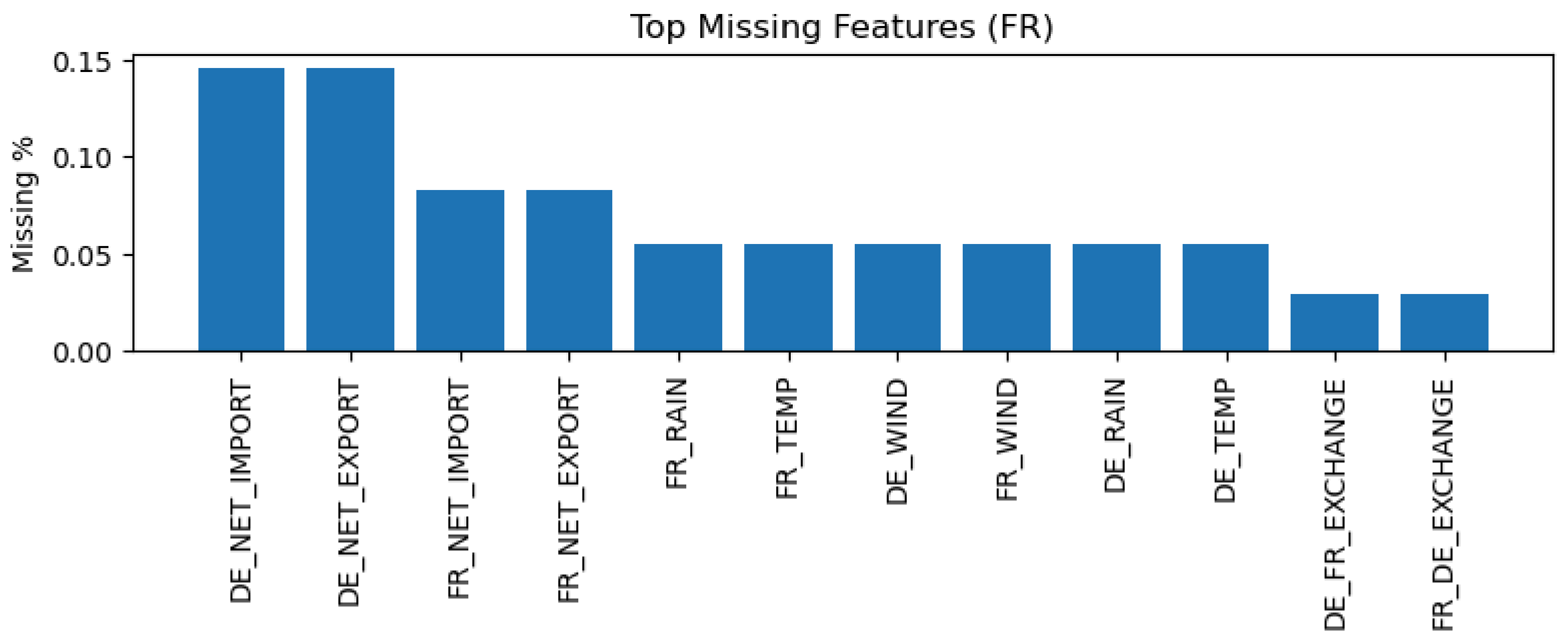
```
        plt.subplot(2, 1, 1)
        topm = mdf.head(miss_count)
        plt.bar(topm.index.astype(str), topm["missing_pct"].values)
        plt.xticks(rotation=90)
        plt.ylabel("Missing %")
        plt.title(f"Top Missing Features ({cc})")
        plt.tight_layout()
        plt.show()

        # Distributions (cap)
        for c in numc0:
            plt.figure()
            plt.subplot(3, 1, 1)
            d[c].plot(kind="hist", bins=40, alpha=0.7)
            plt.xlabel(c)
            plt.title(f"Distribution of {c} ({cc})")
            plt.tight_layout()
            plt.show()

        # Correlation heatmap (within country)
        if len(numc) >= 2:
            corr_c = d[numc].corr()
            plt.figure()
            plt.imshow(corr_c.values, aspect='auto', interpolation='nearest')
            plt.colorbar(label="Pearson r")
            plt.title(f"Feature-Feature Correlation Heatmap ({cc})")
            plt.tight_layout()
            plt.show()

        # Feature-Target correlation (within country)
        if "TARGET" in d.columns and len(numc) > 0:
            tc = d[numc + ["TARGET"]].corr()["TARGET"].drop("TARGET").sort_values(ascending=False)
            display(tc)
            tca = tc.abs().sort_values(ascending=False)
            plt.figure()
            plt.bar(tca.index.astype(str), tca.values)
            plt.xticks(rotation=90)
            plt.ylabel("|Pearson r|")
            plt.title(f"Top Correlation with TARGET ({cc})")
```

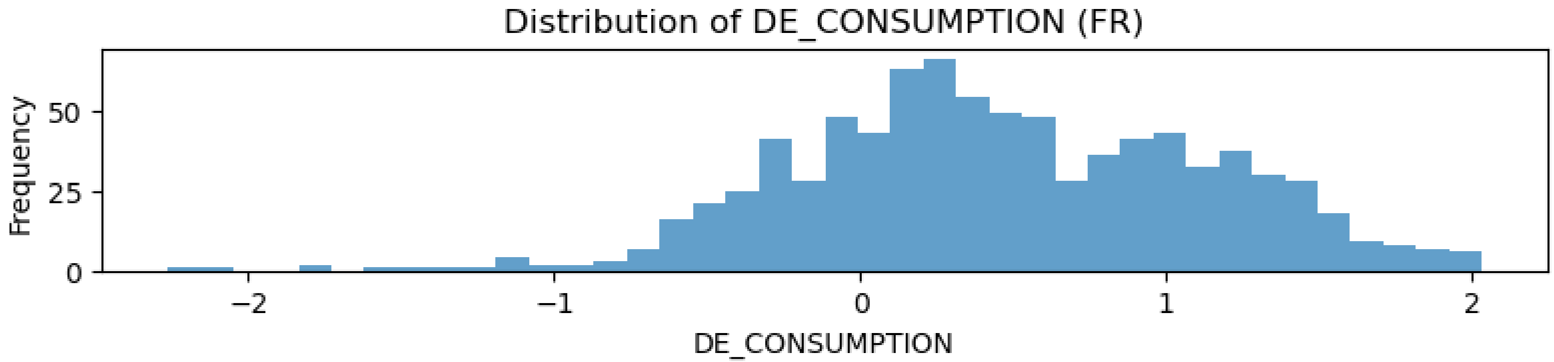
```
        plt.tight_layout()
        plt.show()
```

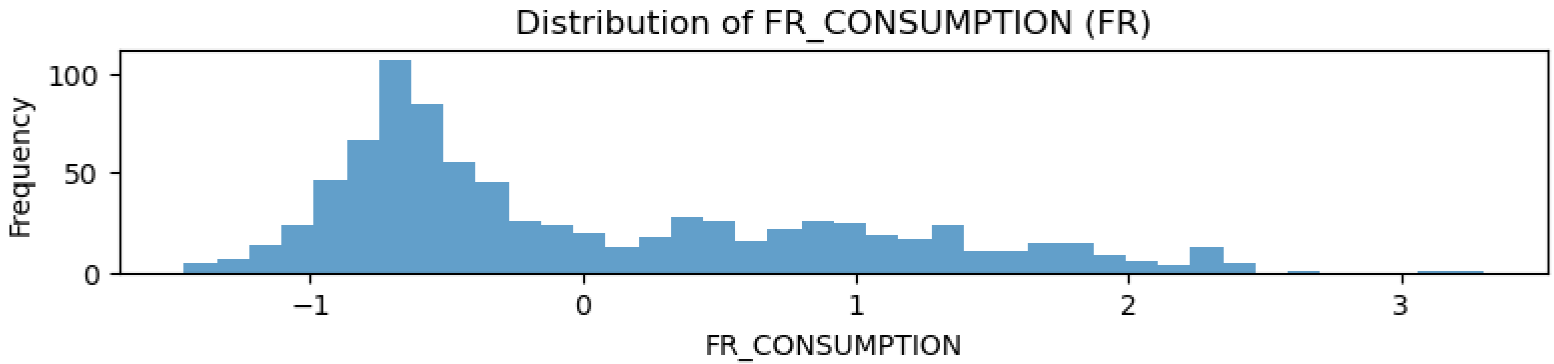
```
Countries found: ['FR', 'DE']

==============================
### COUNTRY: FR
Rows: 851; Numeric features: 33
```

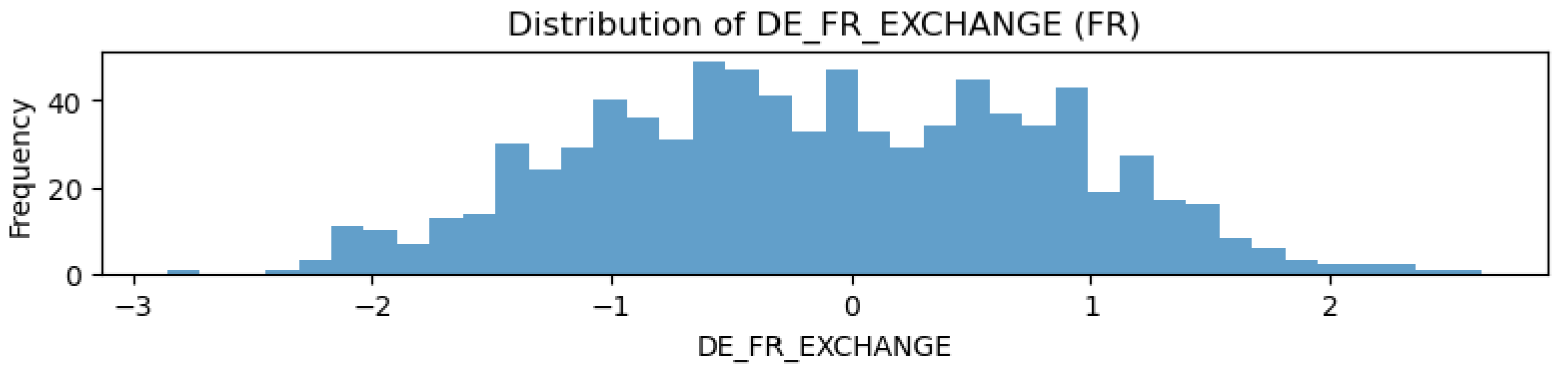
	missing_count	missing_pct
DE_NET_IMPORT	124	0.145711
DE_NET_EXPORT	124	0.145711
FR_NET_IMPORT	70	0.082256
FR_NET_EXPORT	70	0.082256
FR_RAIN	47	0.055229
FR_TEMP	47	0.055229
DE_WIND	47	0.055229
FR_WIND	47	0.055229
DE_RAIN	47	0.055229
DE_TEMP	47	0.055229
DE_FR_EXCHANGE	25	0.029377
FR_DE_EXCHANGE	25	0.029377
DAY_ID	0	0.000000
DE_GAS	0	0.000000
FR_GAS	0	0.000000
DE_CONSUMPTION	0	0.000000
FR_CONSUMPTION	0	0.000000
FR_NUCLEAR	0	0.000000
DE_NUCLEAR	0	0.000000
FR_HYDRO	0	0.000000
DE_HYDRO	0	0.000000
FR_COAL	0	0.000000

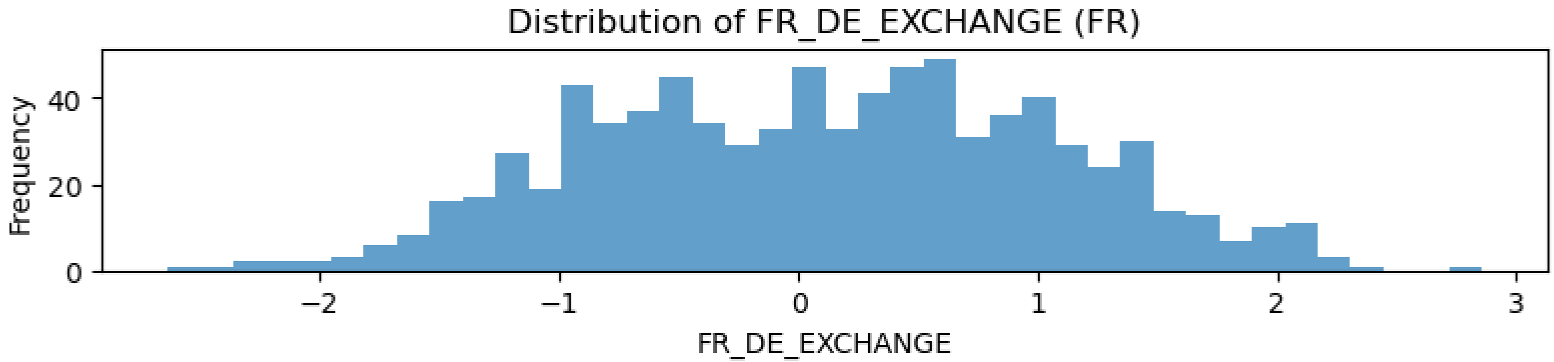
	missing_count	missing_pct
DE_COAL	0	0.000000
DE_WINDPOW	0	0.000000
DE_SOLAR	0	0.000000
FR_RESIDUAL_LOAD	0	0.000000
DE_RESIDUAL_LOAD	0	0.000000
DE_LIGNITE	0	0.000000
FR_WINDPOW	0	0.000000
FR_SOLAR	0	0.000000
GAS_RET	0	0.000000
COAL_RET	0	0.000000
CARBON_RET	0	0.000000

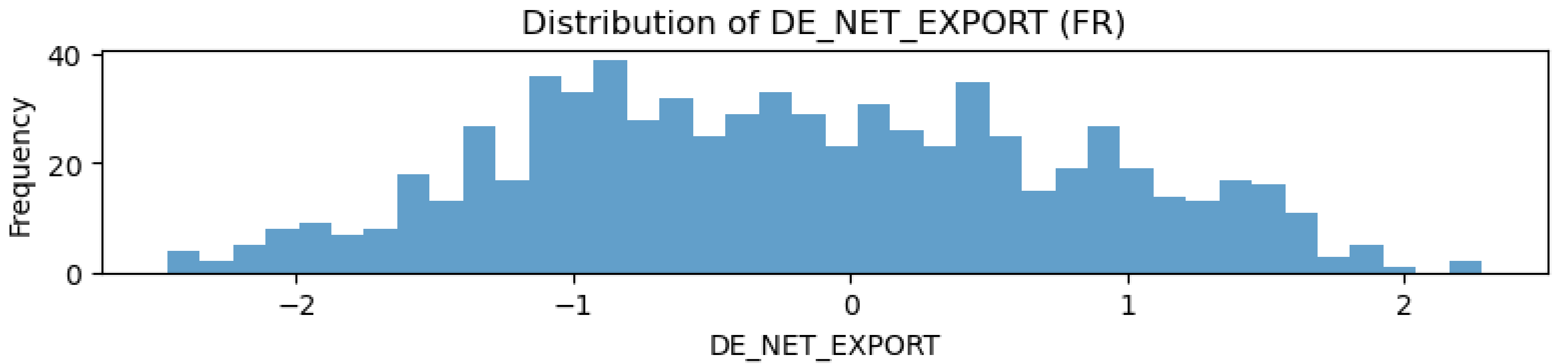


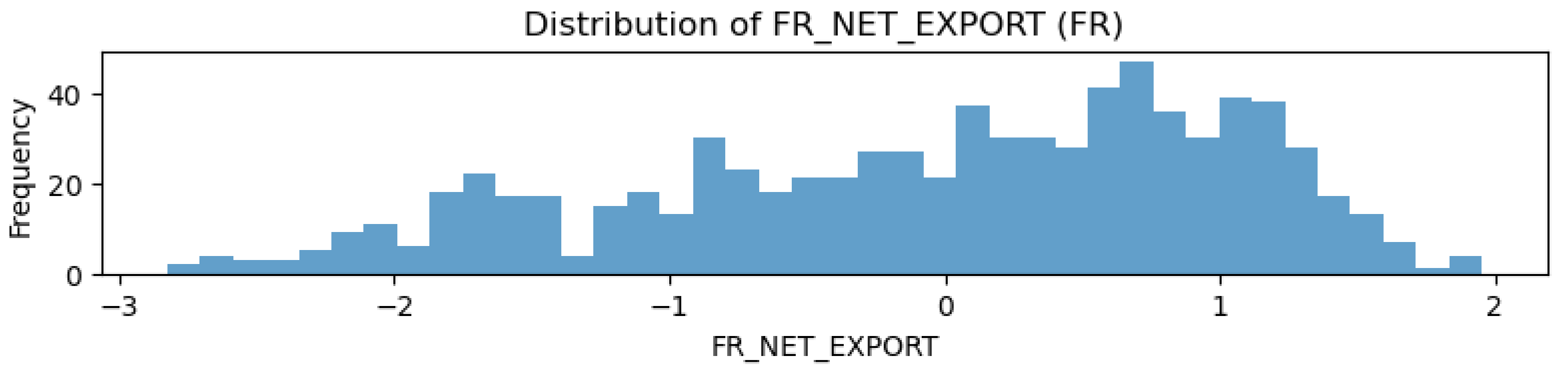

Distribution of DE_CONSUMPTION (FR)
Frequency
0
25
50
−2
−1
0
1
2
DE_CONSUMPTION

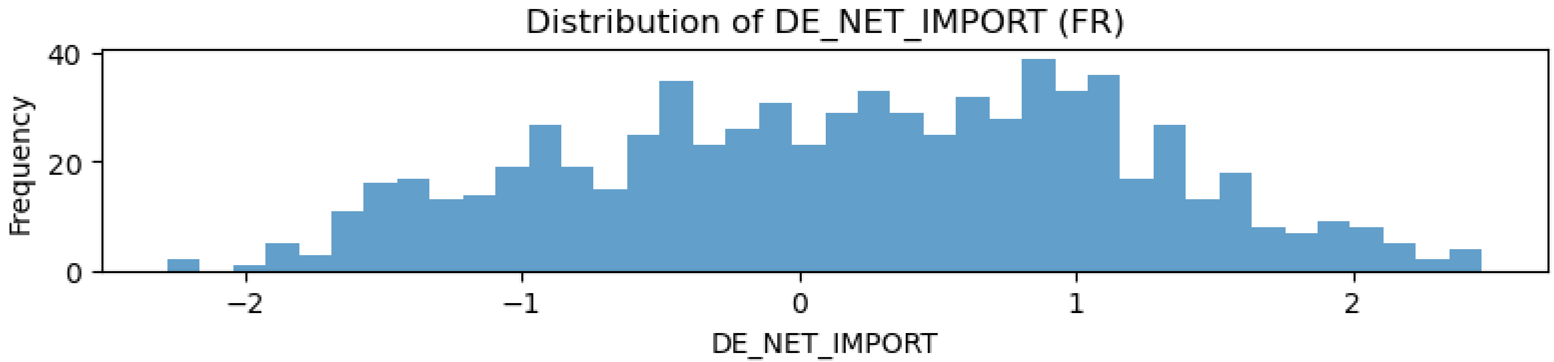

Distribution of FR_CONSUMPTION (FR)
Frequency
0
50
100
−1
0
1
2
3
FR_CONSUMPTION

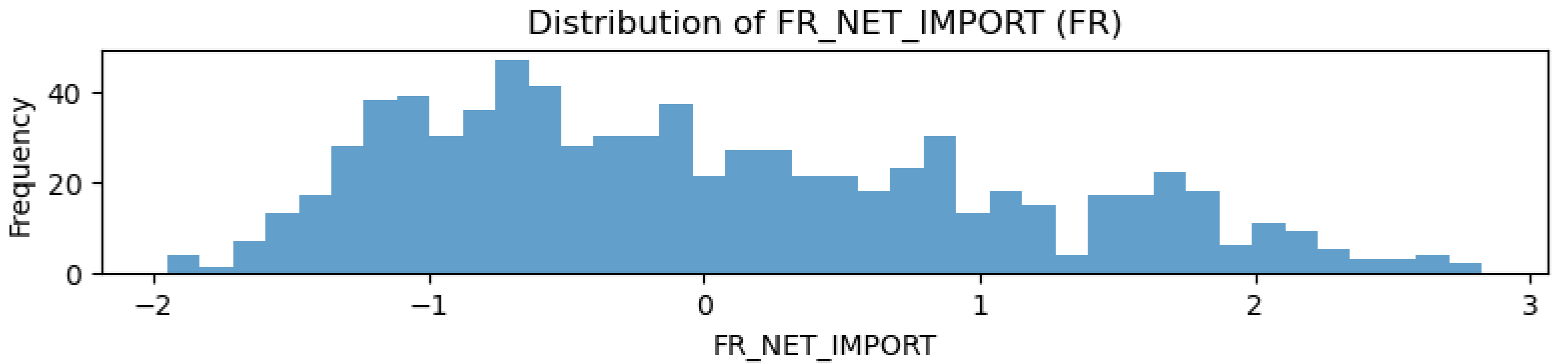

Distribution of DE_FR_EXCHANGE (FR)
Frequency
0
20
40
−3
−2
−1
0
1
2
DE_FR_EXCHANGE

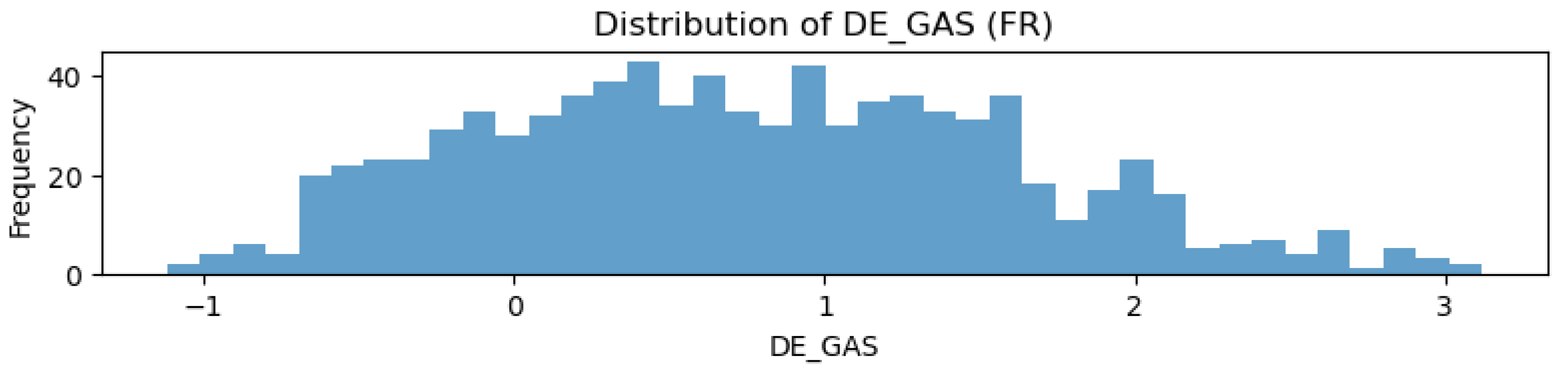

Distribution of FR_DE_EXCHANGE (FR)
Frequency
0
20
40
−2
−1
0
1
2
3
FR_DE_EXCHANGE

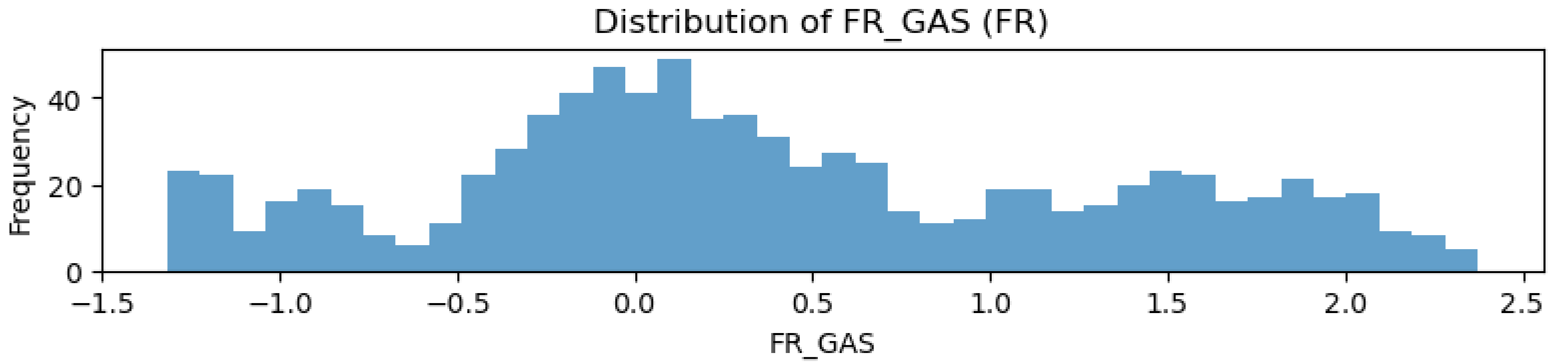

Distribution of DE_NET_EXPORT (FR)
Frequency
0
20
40
−2
−1
0
1
2
DE_NET_EXPORT

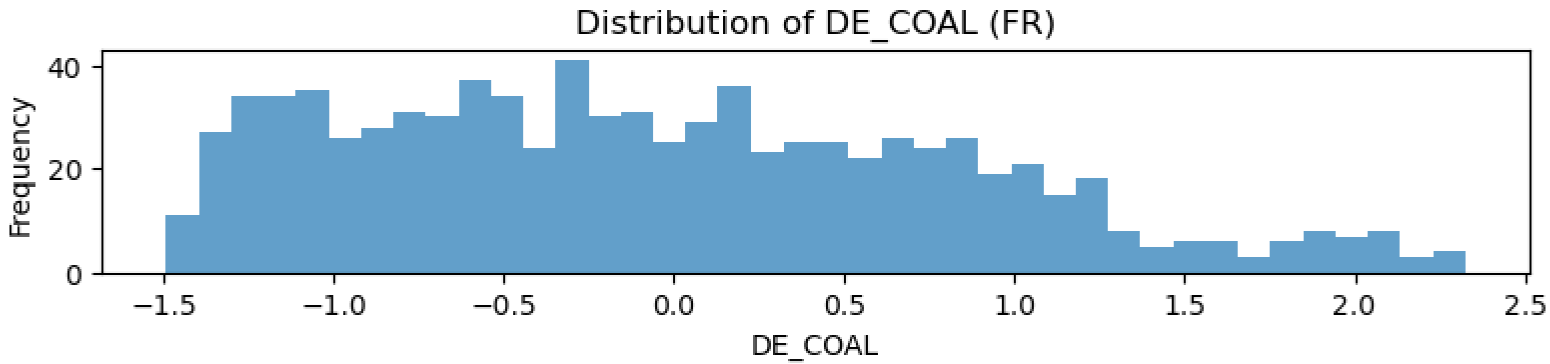

Distribution of FR_NET_EXPORT (FR)
Frequency
0
20
40
−3
−2
−1
0
1
2
FR_NET_EXPORT

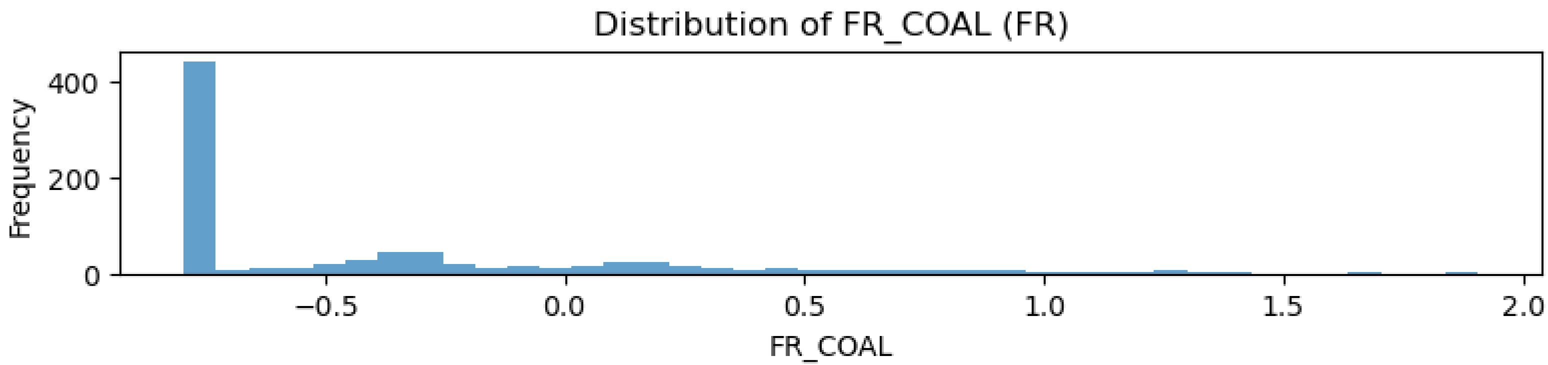

Distribution of DE_NET_IMPORT (FR)
Frequency
0
20
40
−2
−1
0
1
2
DE_NET_IMPORT

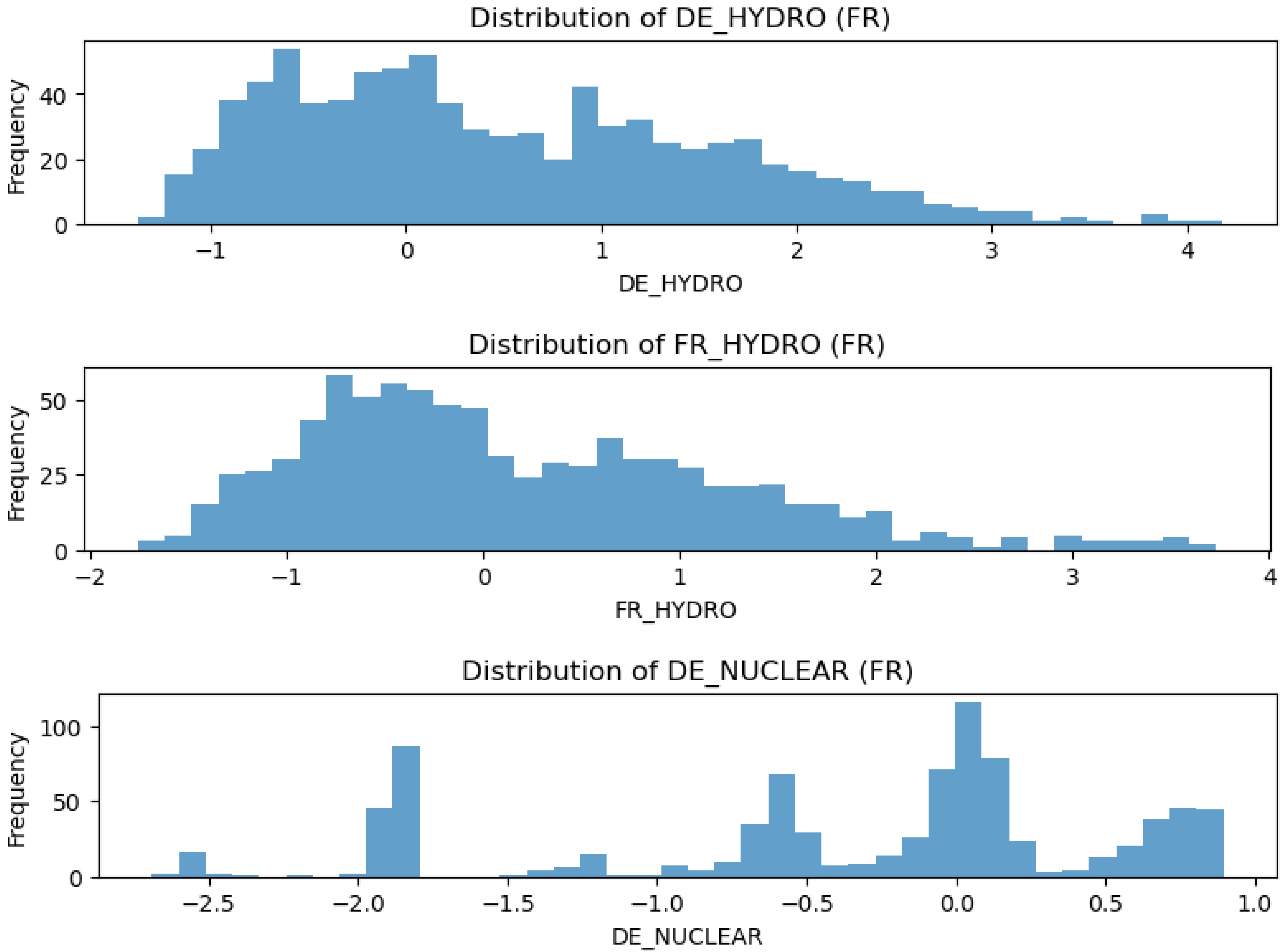

Distribution of FR_NET_IMPORT (FR)
Frequency
0
20
40
−2
−1
0
1
2
3
FR_NET_IMPORT

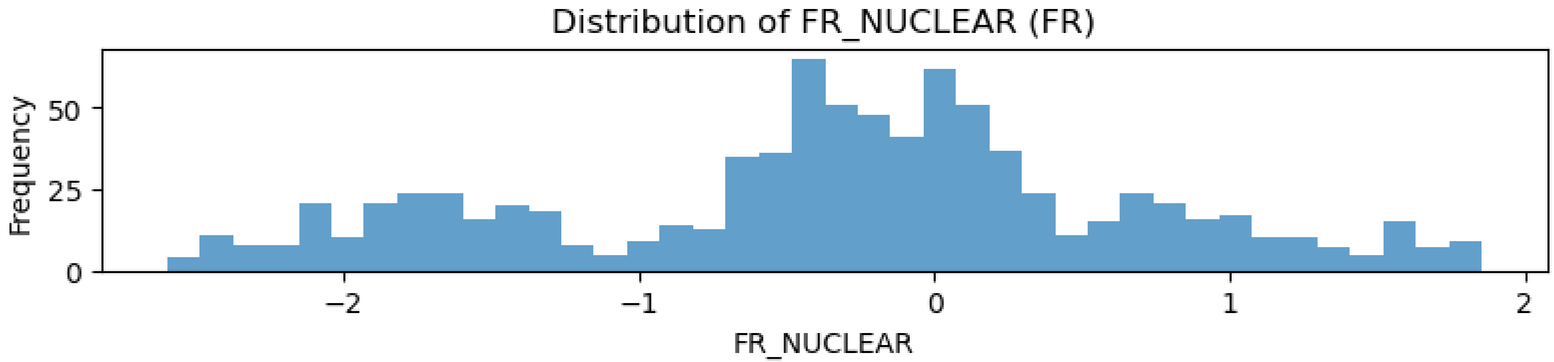

Distribution of DE_GAS (FR)
Frequency
0
20
40
−1
0
1
2
3
DE_GAS

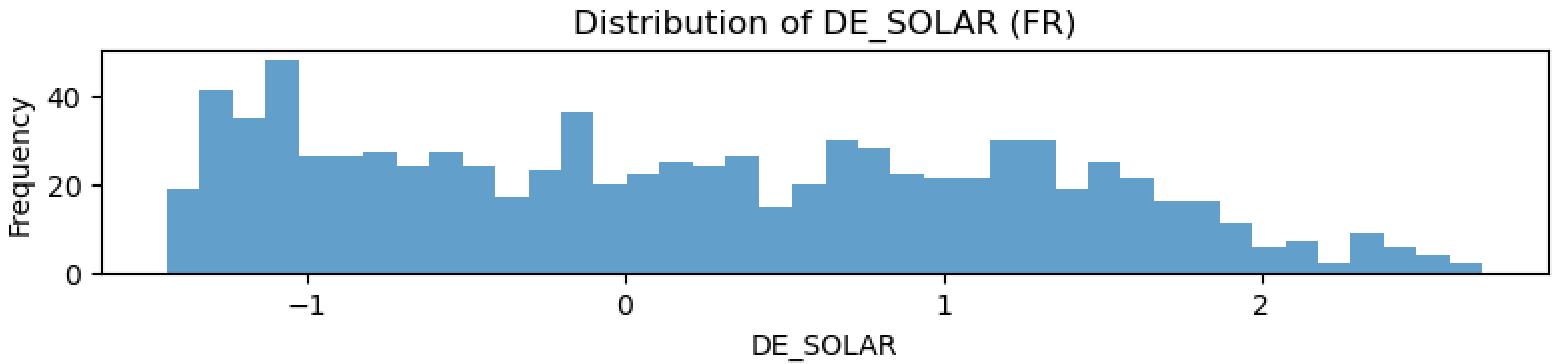

Distribution of FR_GAS (FR)
Frequency
0
20
40
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
FR_GAS

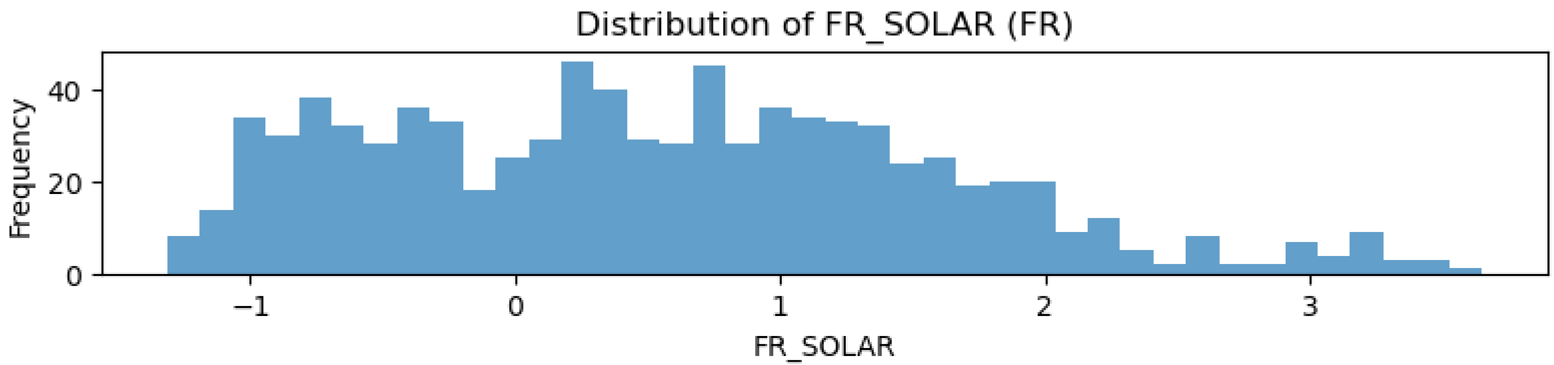

Distribution of DE_COAL (FR)
Frequency
0
20
40
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
DE_COAL

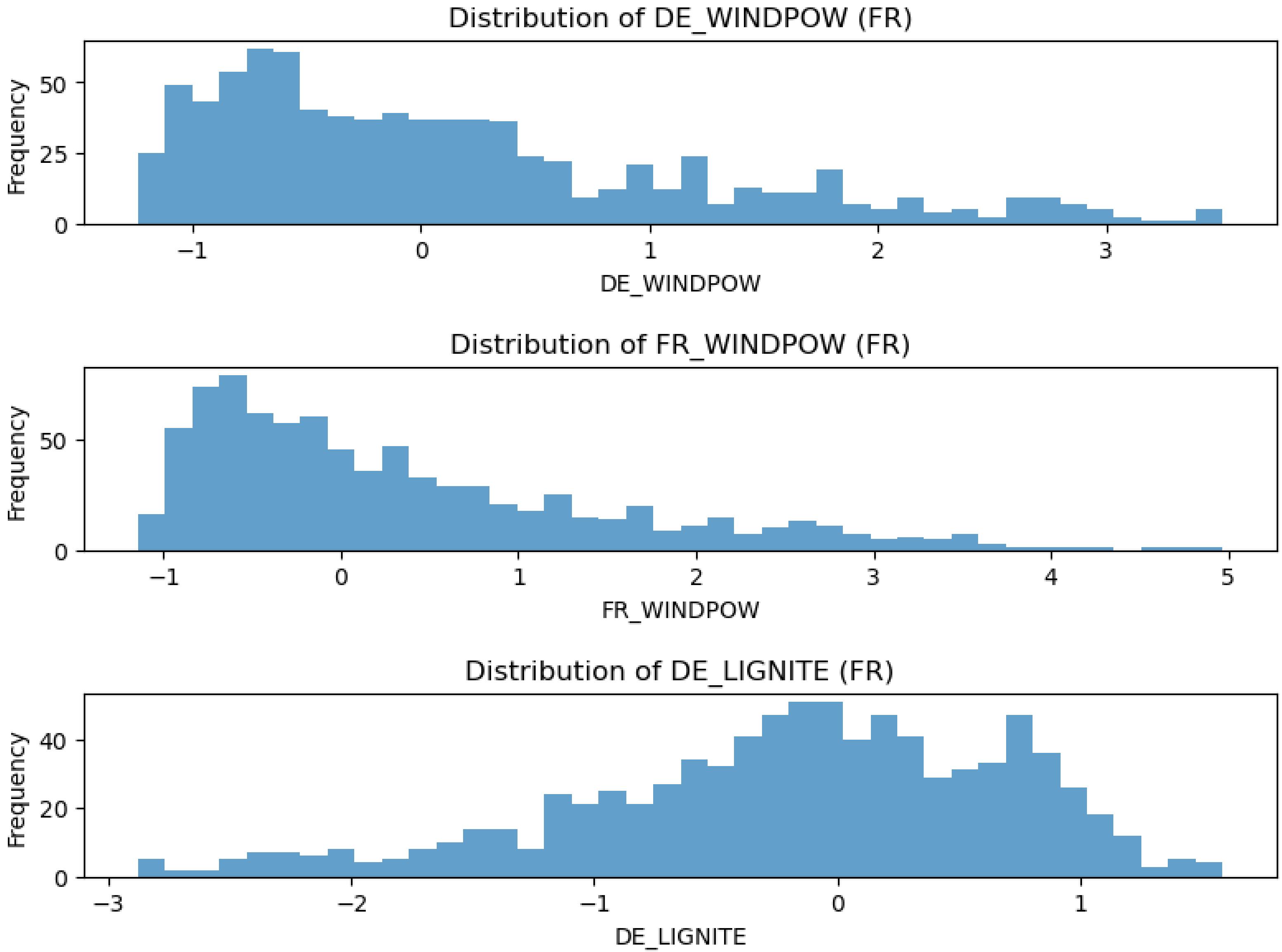

Distribution of FR_COAL (FR)
Frequency
0
200
400
−0.5
0.0
0.5
1.0
1.5
2.0
FR_COAL

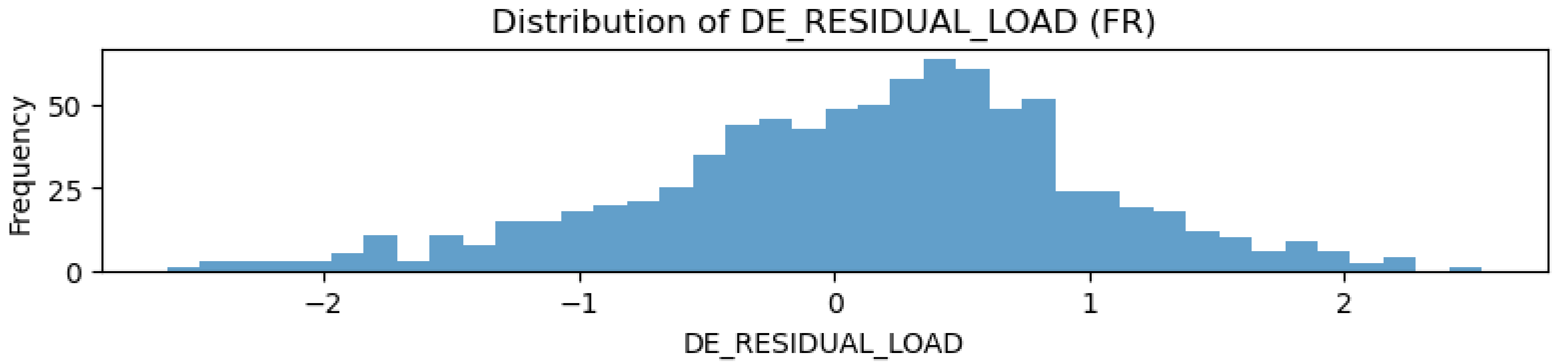

Distribution of DE_HYDRO (FR)
Frequency
0
20
40
−1
0
1
2
3
4
DE_HYDRO
Distribution of FR_HYDRO (FR)
Frequency
0
25
50
−2
−1
0
1
2
3
4
FR_HYDRO
Distribution of DE_NUCLEAR (FR)
Frequency
0
50
100
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
DE_NUCLEAR

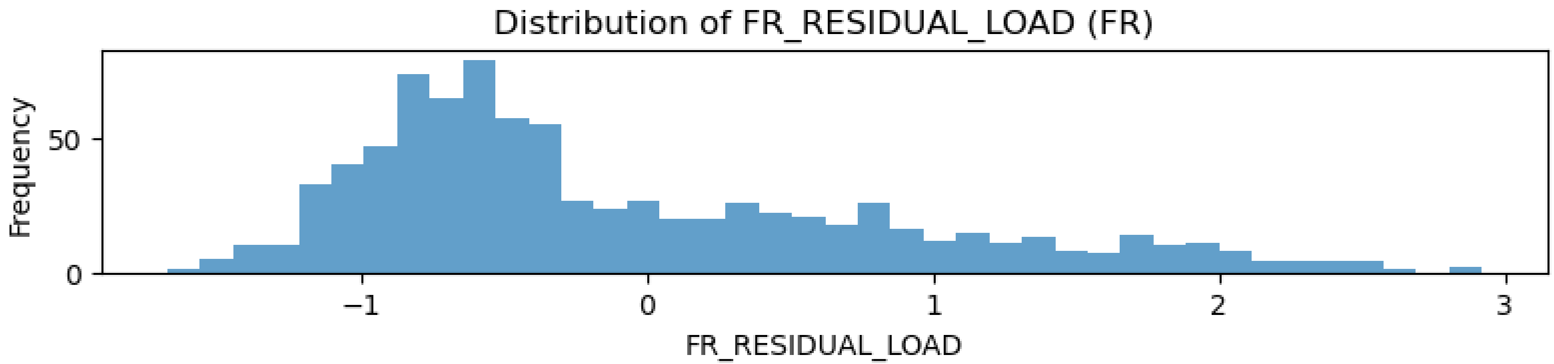

Distribution of FR_NUCLEAR (FR)
Frequency
0
25
50
−2
−1
0
1
2
FR_NUCLEAR

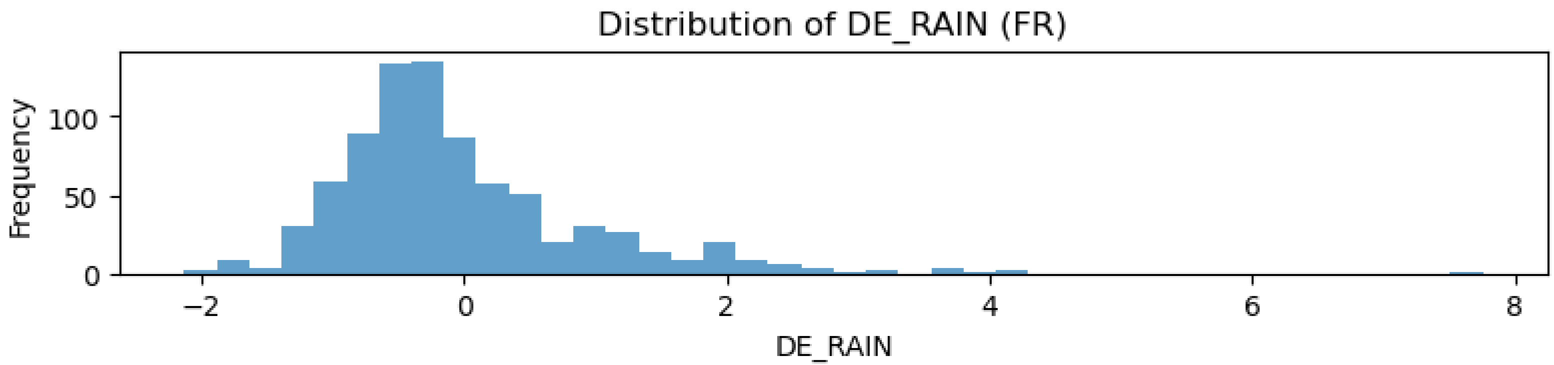

Distribution of DE_SOLAR (FR)
Frequency
0
20
40
−1
0
1
2
DE_SOLAR

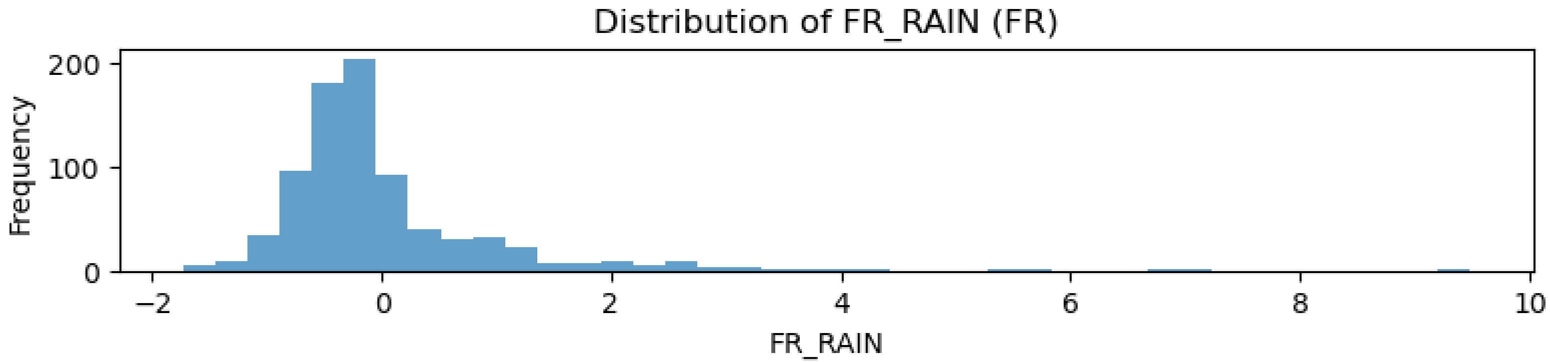

Distribution of FR_SOLAR (FR)
Frequency
0
20
40
−1
0
1
2
3
FR_SOLAR

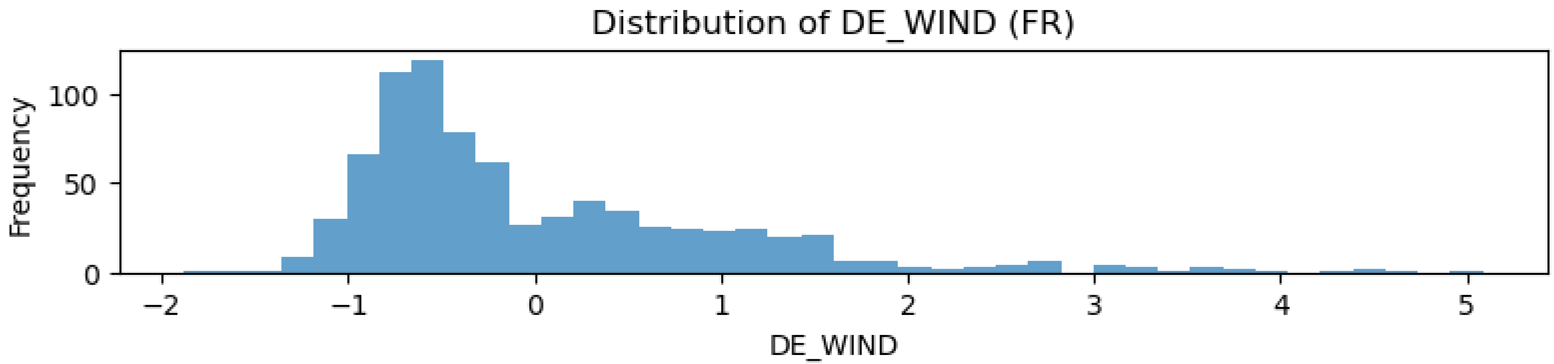

Distribution of DE_WINDPOW (FR)
Frequency
0
25
50
−1
0
1
2
3
DE_WINDPOW
Distribution of FR_WINDPOW (FR)
Frequency
0
50
−1
0
1
2
3
4
5
FR_WINDPOW
Distribution of DE_LIGNITE (FR)
Frequency
0
20
40
−3
−2
−1
0
1
DE_LIGNITE

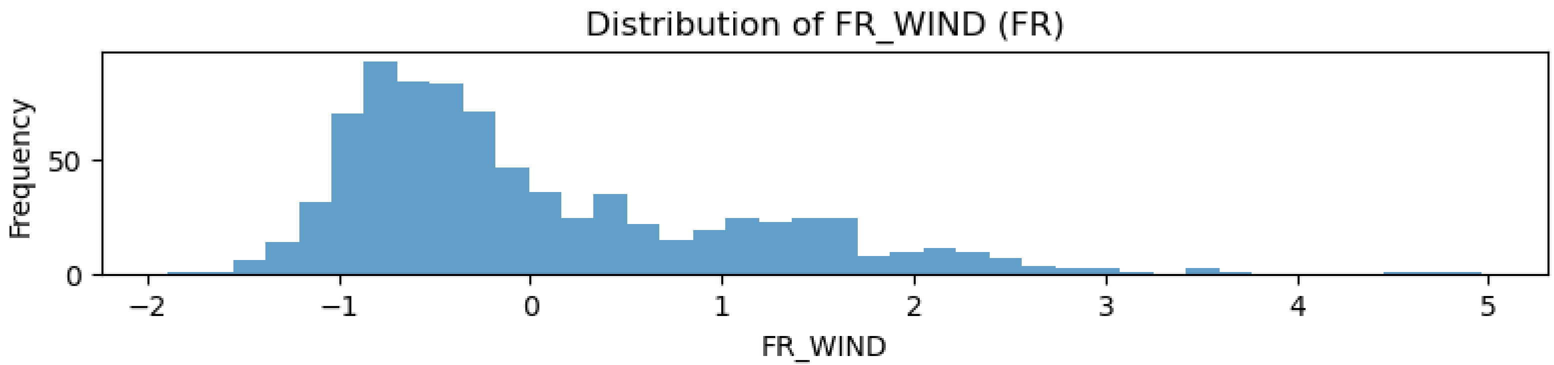

Distribution of DE_RESIDUAL_LOAD (FR)
Frequency
0
25
50
−2
−1
0
1
2
DE_RESIDUAL_LOAD

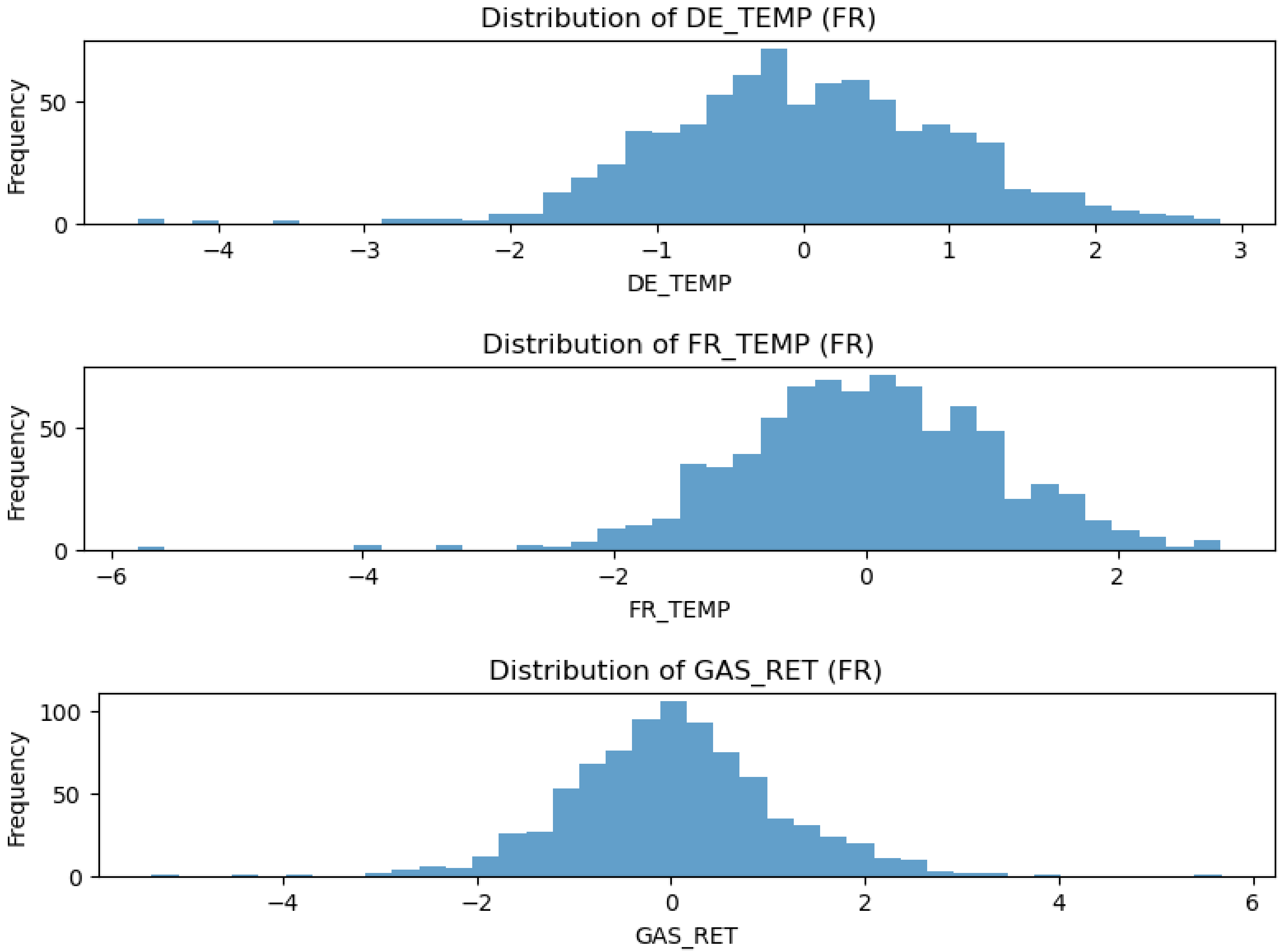

Distribution of FR_RESIDUAL_LOAD (FR)
Frequency
0
50
−1
0
1
2
3
FR_RESIDUAL_LOAD

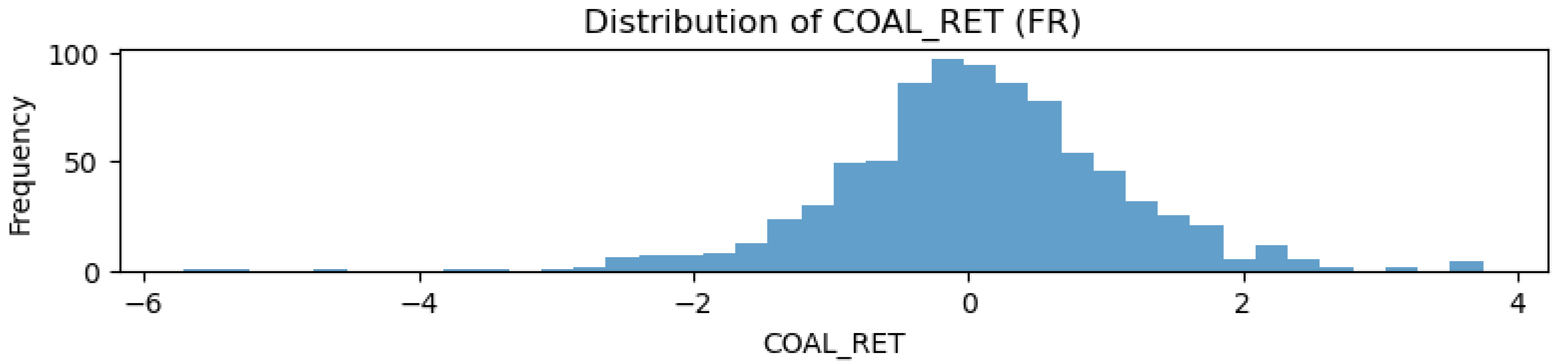

Distribution of DE_RAIN (FR)
Frequency
0
50
100
−2
0
2
4
6
8
DE_RAIN

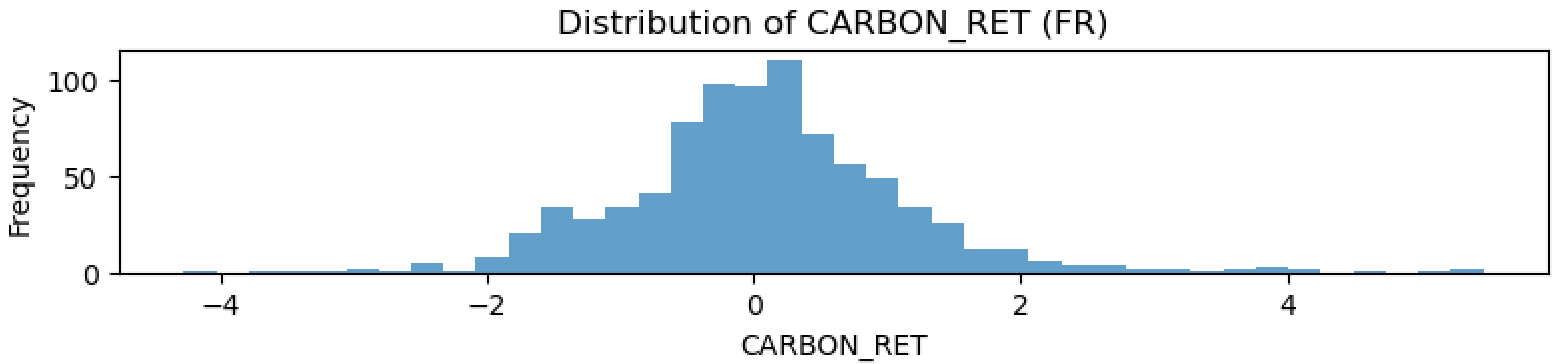

Distribution of FR_RAIN (FR)
Frequency
200
100
0
−2
0
2
4
6
8
10
FR_RAIN

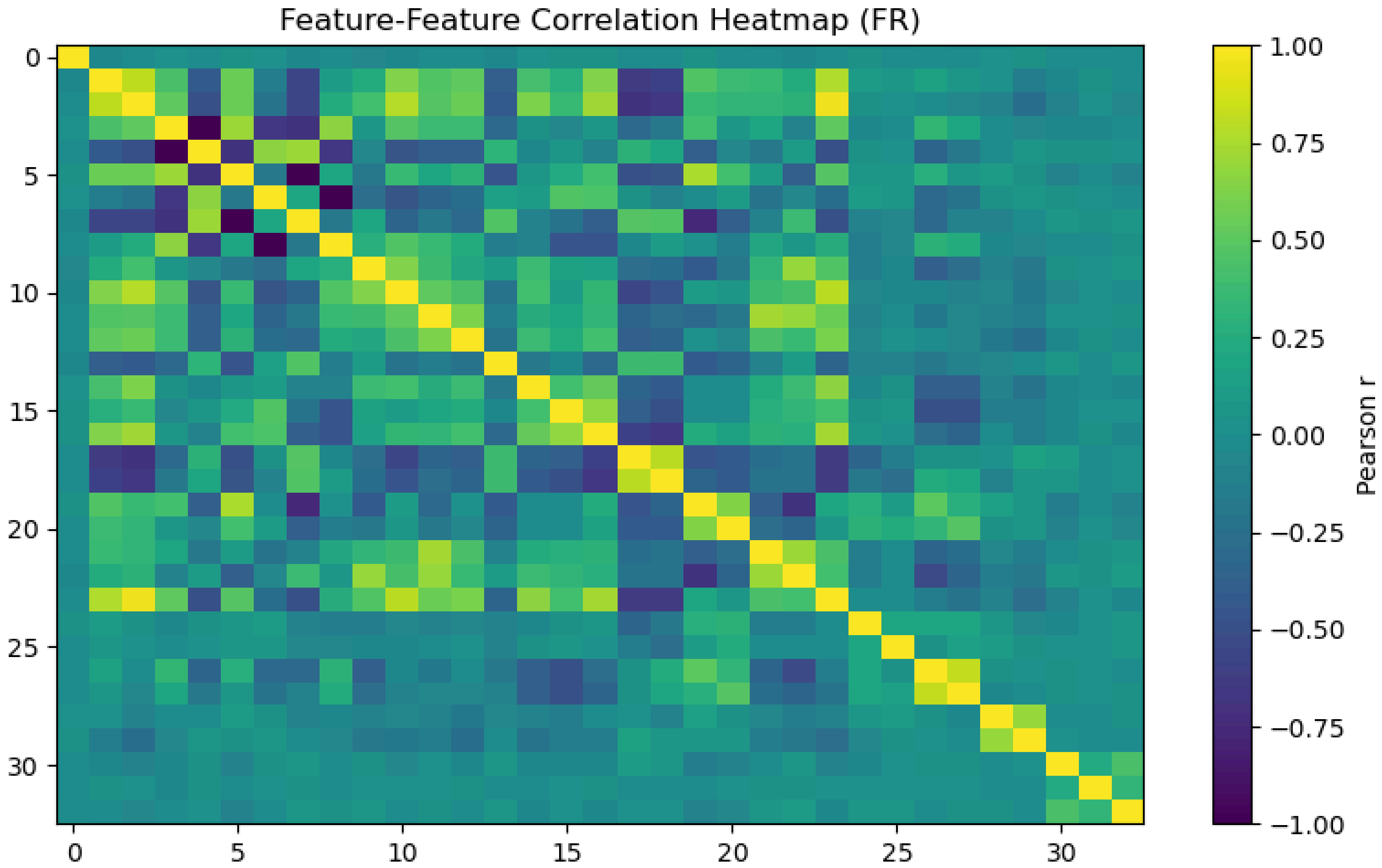

Distribution of DE_WIND (FR)
Frequency
100
50
0
−2
−1
0
1
2
3
4
5
DE_WIND


Distribution of FR_WIND (FR)
Frequency
50
0
−2
−1
0
1
2
3
4
5
FR_WIND

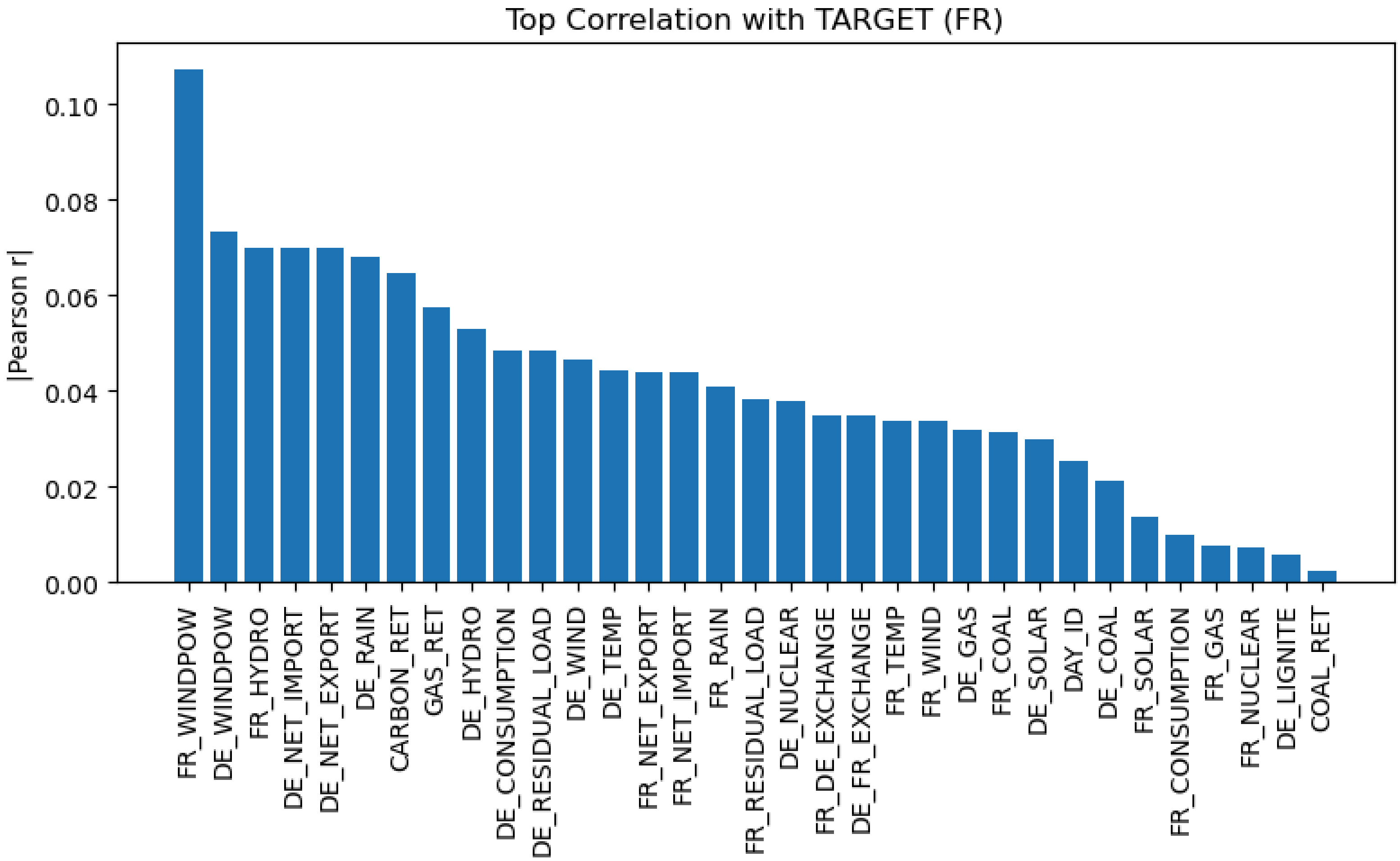

Distribution of DE_TEMP (FR)
Frequency
50
0
−4
−3
−2
−1
0
1
2
3
DE_TEMP
Distribution of FR_TEMP (FR)
Frequency
50
0
−6
−4
−2
0
2
FR_TEMP
Distribution of GAS_RET (FR)
Frequency
100
50
0
−4
−2
0
2
4
6
GAS_RET


Distribution of COAL_RET (FR)
Frequency
100
50
0
−6
−4
−2
0
2
4
COAL_RET


Distribution of CARBON_RET (FR)
Frequency
100
50
0
−4
−2
0
2
4
CARBON_RET


Feature-Feature Correlation Heatmap (FR)
0
5
10
15
20
25
30
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Pearson r

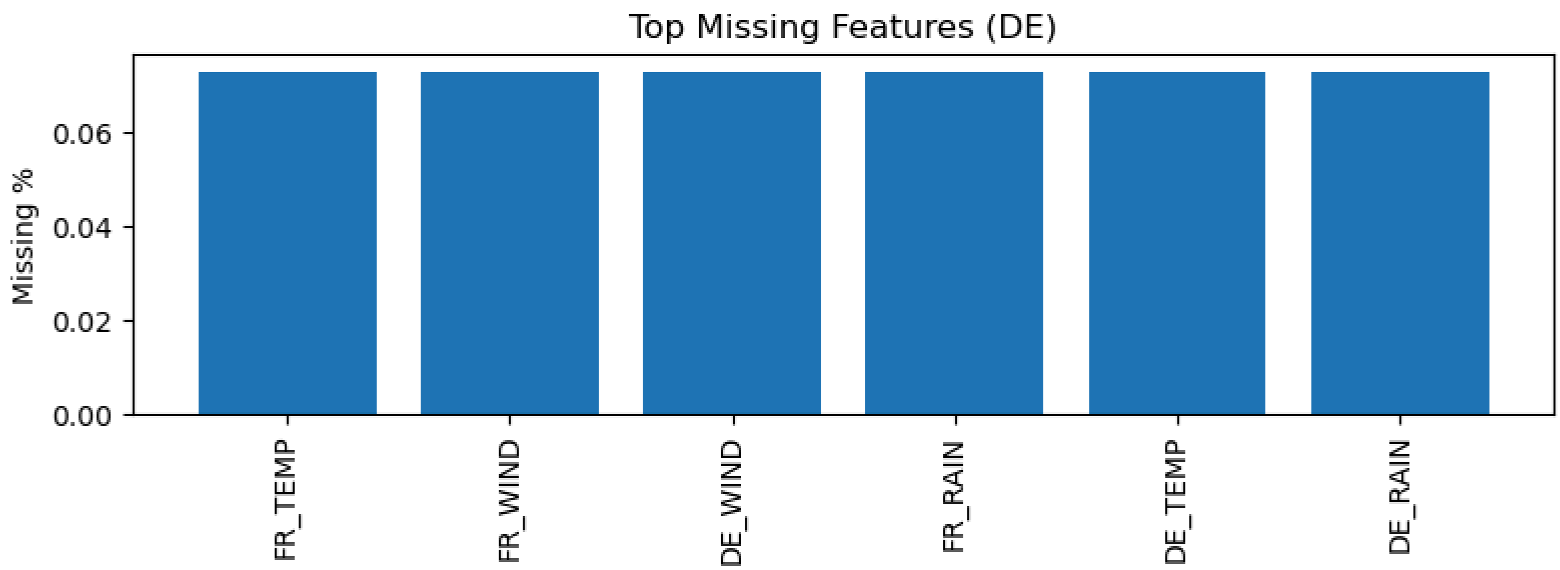
```
FR_HYDRO            0.070002
DE_NET_IMPORT       0.069824
CARBON_RET          0.064776
GAS_RET             0.057536
DE_HYDRO            0.052827
DE_RESIDUAL_LOAD    0.048185
FR_NET_IMPORT       0.043865
FR_RESIDUAL_LOAD    0.038176
DE_NUCLEAR          0.037859
DE_FR_EXCHANGE      0.034964
DE_GAS              0.031626
FR_COAL             0.031529
DE_SOLAR            0.029904
DE_COAL             0.021014
FR_SOLAR            0.013828
FR_CONSUMPTION      0.009998
FR_GAS              0.007433
FR_NUCLEAR          0.007304
DE_LIGNITE          0.005673
COAL_RET           -0.002330
DAY_ID             -0.025336
FR_WIND            -0.033683
FR_TEMP            -0.033776
FR_DE_EXCHANGE     -0.034964
FR_RAIN            -0.040963
FR_NET_EXPORT      -0.043865
DE_TEMP            -0.044049
DE_WIND            -0.046644
DE_CONSUMPTION     -0.048551
DE_RAIN            -0.067974
DE_NET_EXPORT      -0.069824
DE_WINDPOW         -0.073254
FR_WINDPOW         -0.107403
Name: TARGET, dtype: float64
```

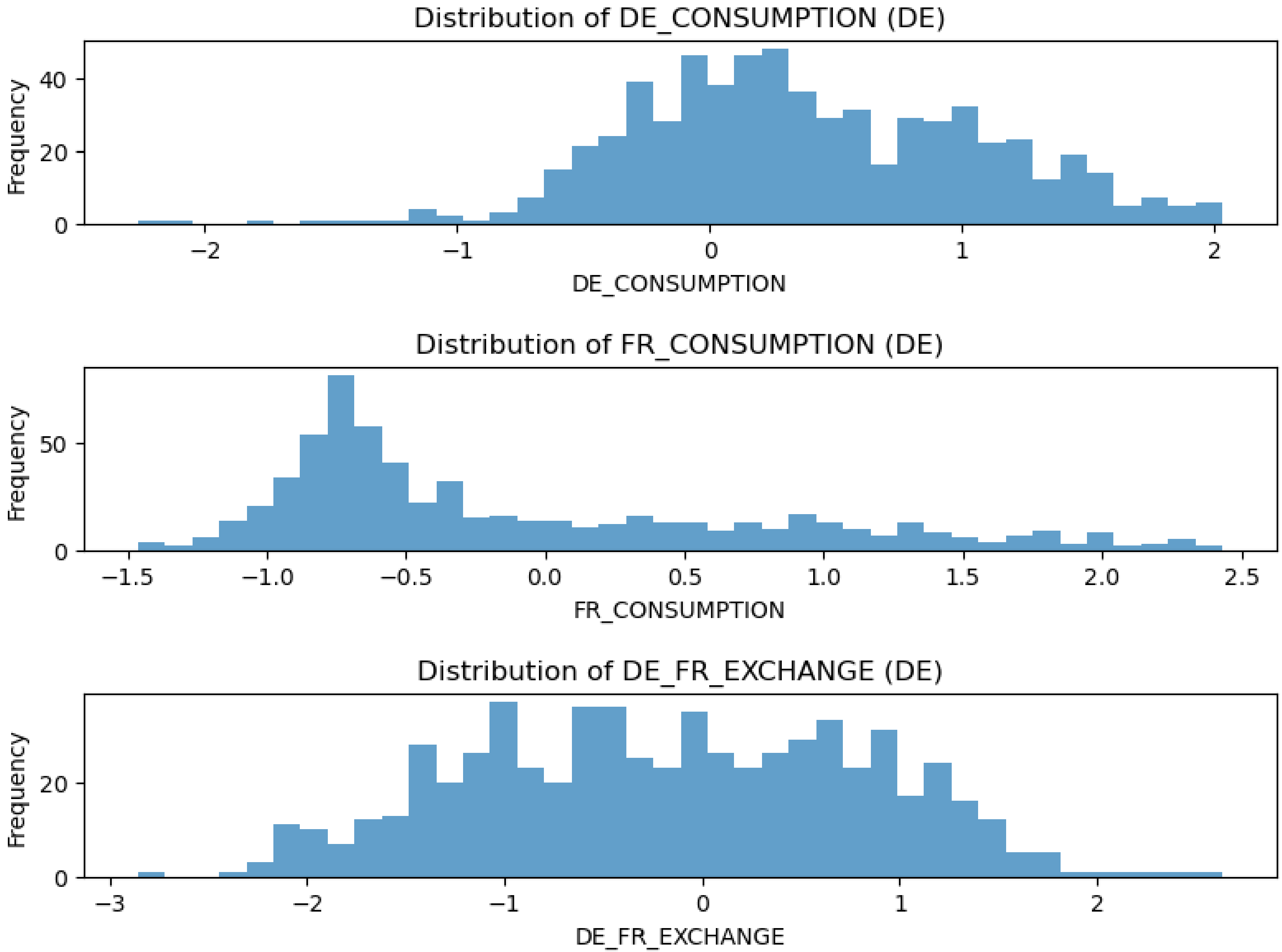


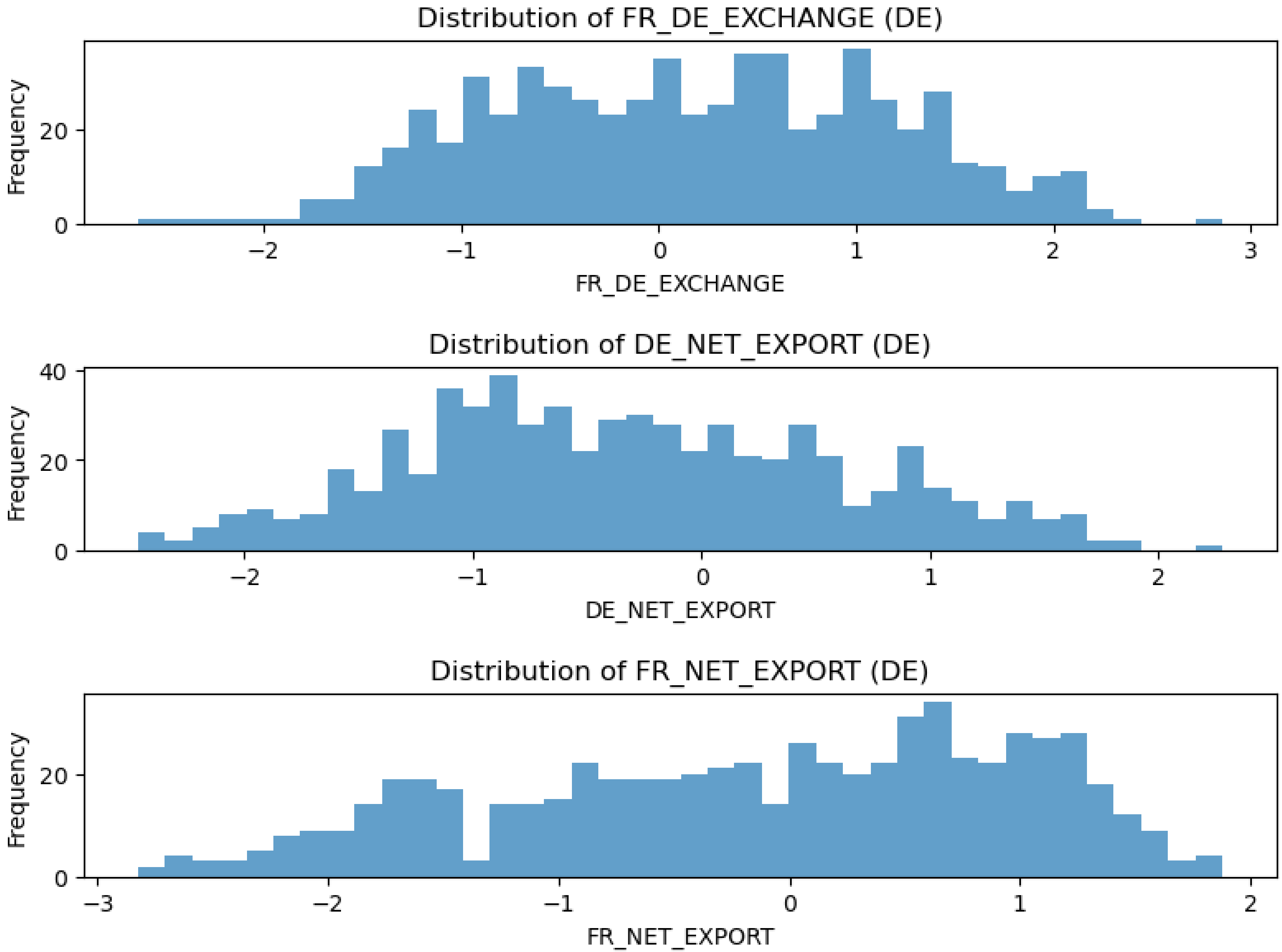

```
==============================
### COUNTRY: DE
Rows: 643; Numeric features: 33
```

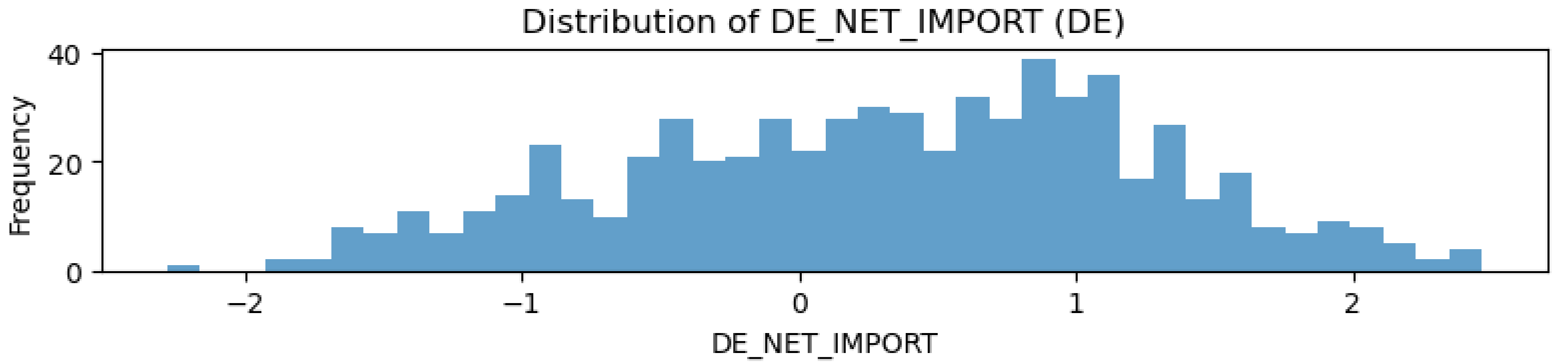
	missing_count	missing_pct
FR_TEMP	47	0.073095
FR_WIND	47	0.073095
DE_WIND	47	0.073095
FR_RAIN	47	0.073095
DE_TEMP	47	0.073095
DE_RAIN	47	0.073095
DAY_ID	0	0.000000
DE_CONSUMPTION	0	0.000000
FR_CONSUMPTION	0	0.000000
DE_GAS	0	0.000000
FR_GAS	0	0.000000
DE_FR_EXCHANGE	0	0.000000
FR_DE_EXCHANGE	0	0.000000
DE_NET_EXPORT	0	0.000000
FR_NET_EXPORT	0	0.000000
DE_NET_IMPORT	0	0.000000
FR_NET_IMPORT	0	0.000000
FR_NUCLEAR	0	0.000000
DE_NUCLEAR	0	0.000000
FR_HYDRO	0	0.000000
DE_HYDRO	0	0.000000
FR_COAL	0	0.000000

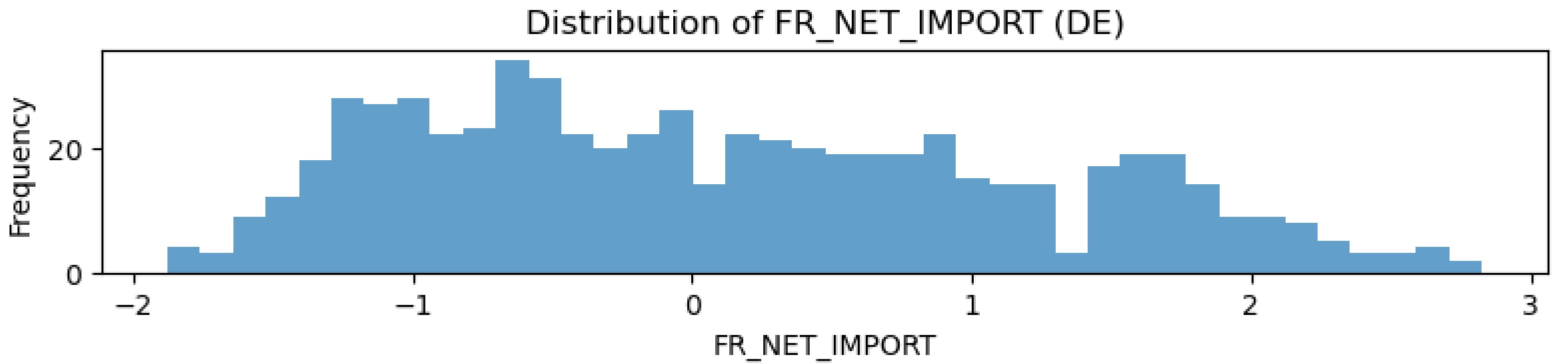
	missing_count	missing_pct
DE_COAL	0	0.000000
DE_WINDPOW	0	0.000000
DE_SOLAR	0	0.000000
FR_RESIDUAL_LOAD	0	0.000000
DE_RESIDUAL_LOAD	0	0.000000
DE_LIGNITE	0	0.000000
FR_WINDPOW	0	0.000000
FR_SOLAR	0	0.000000
GAS_RET	0	0.000000
COAL_RET	0	0.000000
CARBON_RET	0	0.000000

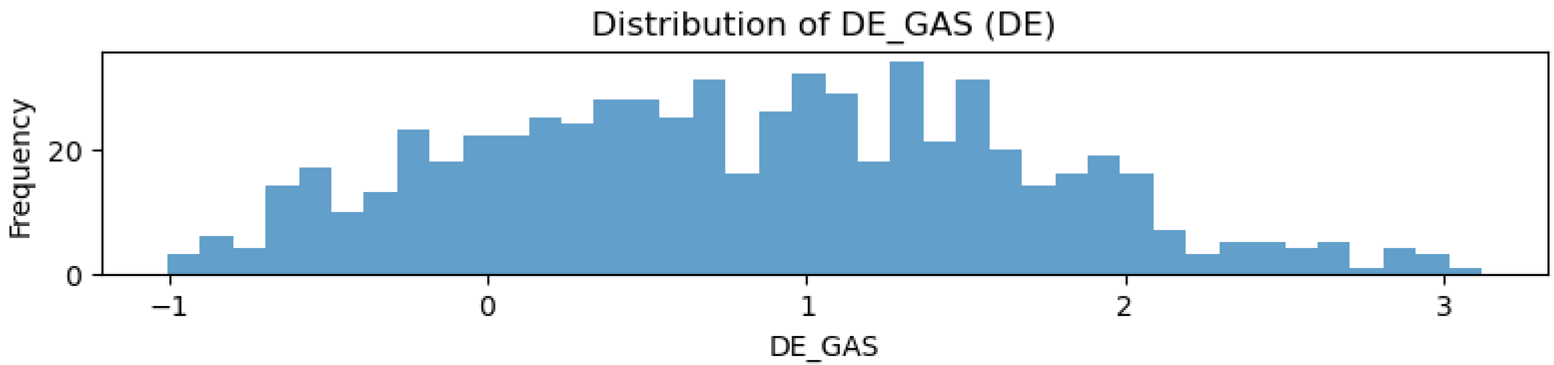


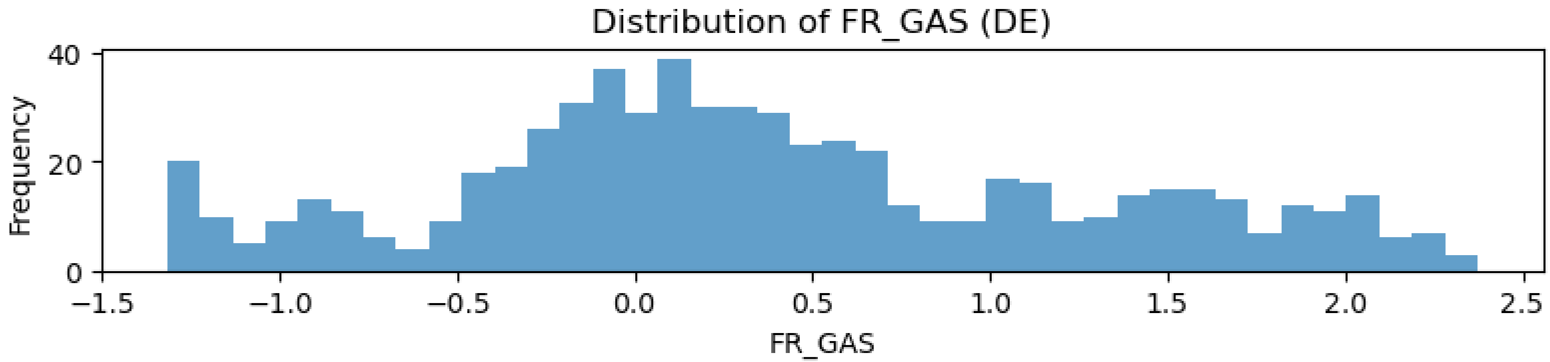

Distribution of DE_CONSUMPTION (DE)
Frequency
0
20
40
−2
−1
0
1
2
DE_CONSUMPTION
Distribution of FR_CONSUMPTION (DE)
Frequency
0
50
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
FR_CONSUMPTION
Distribution of DE_FR_EXCHANGE (DE)
Frequency
0
20
−3
−2
−1
0
1
2
DE_FR_EXCHANGE


Distribution of FR_DE_EXCHANGE (DE)
Frequency
0
20
−2
−1
0
1
2
3
FR_DE_EXCHANGE
Distribution of DE_NET_EXPORT (DE)
Frequency
0
20
40
−2
−1
0
1
2
DE_NET_EXPORT
Distribution of FR_NET_EXPORT (DE)
Frequency
0
20
−3
−2
−1
0
1
2
FR_NET_EXPORT

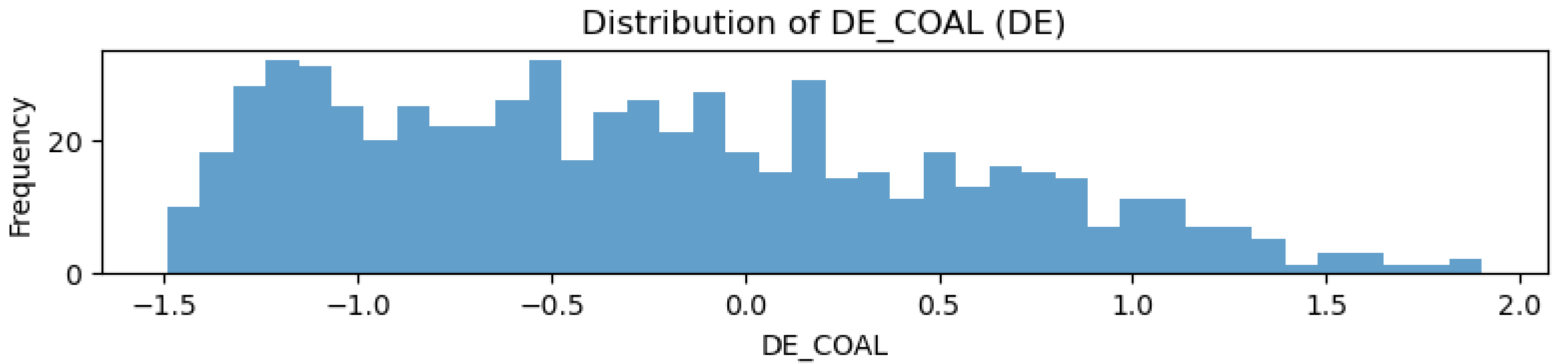

Distribution of DE_NET_IMPORT (DE)
Frequency
40
20
0
−2
−1
0
1
2
DE_NET_IMPORT


Distribution of FR_NET_IMPORT (DE)
Frequency
20
0
−2
−1
0
1
2
3
FR_NET_IMPORT

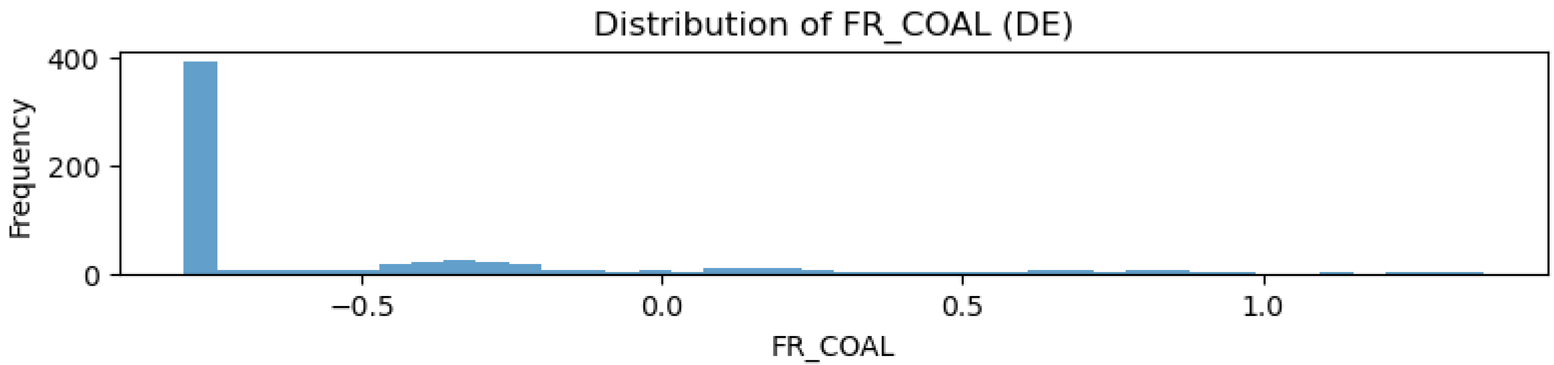

Distribution of DE_GAS (DE)
Frequency
20
0
−1
0
1
2
3
DE_GAS



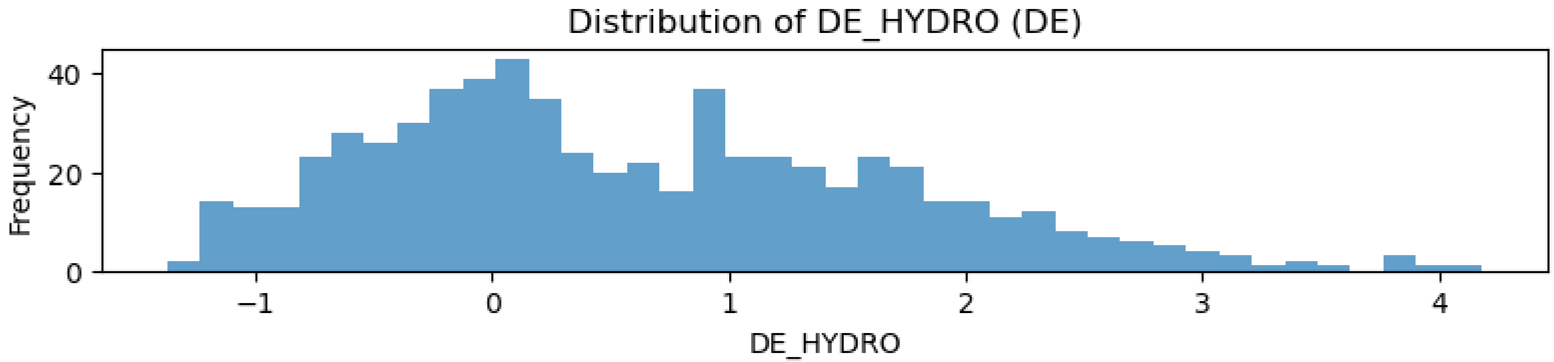
Distribution of FR_GAS (DE)
Frequency
0
20
40
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
FR_GAS

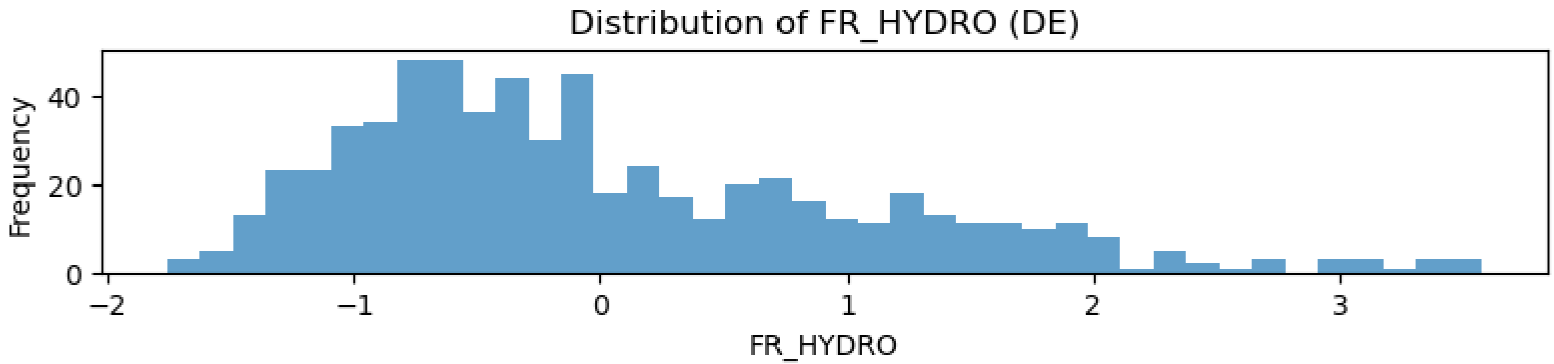


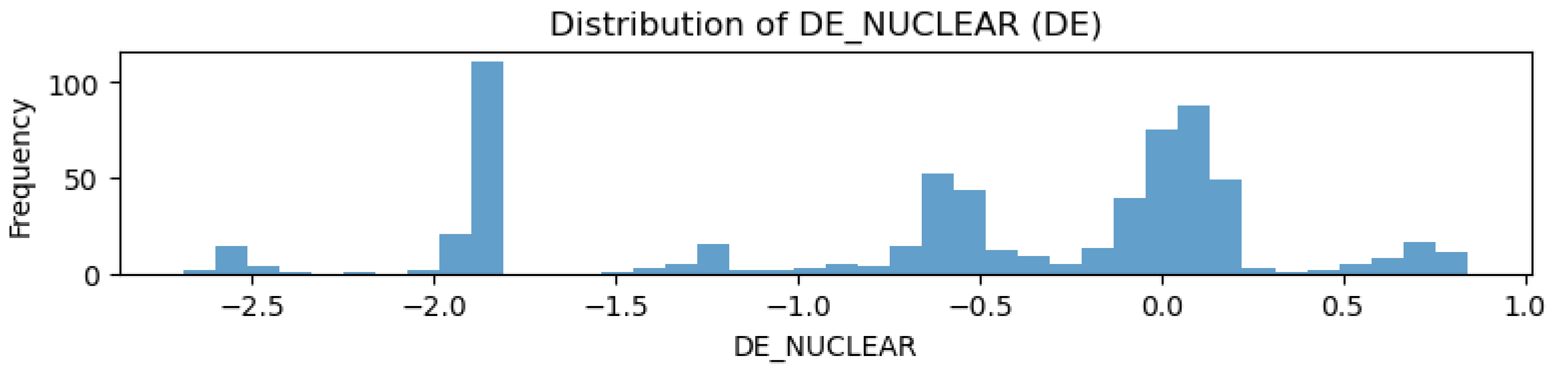
Distribution of DE_COAL (DE)
Frequency
0
20
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
DE_COAL

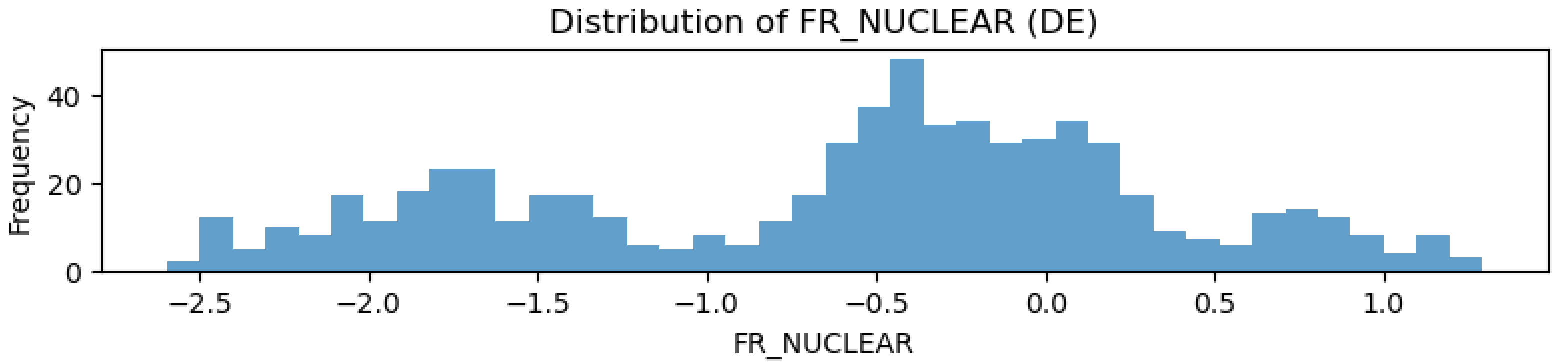


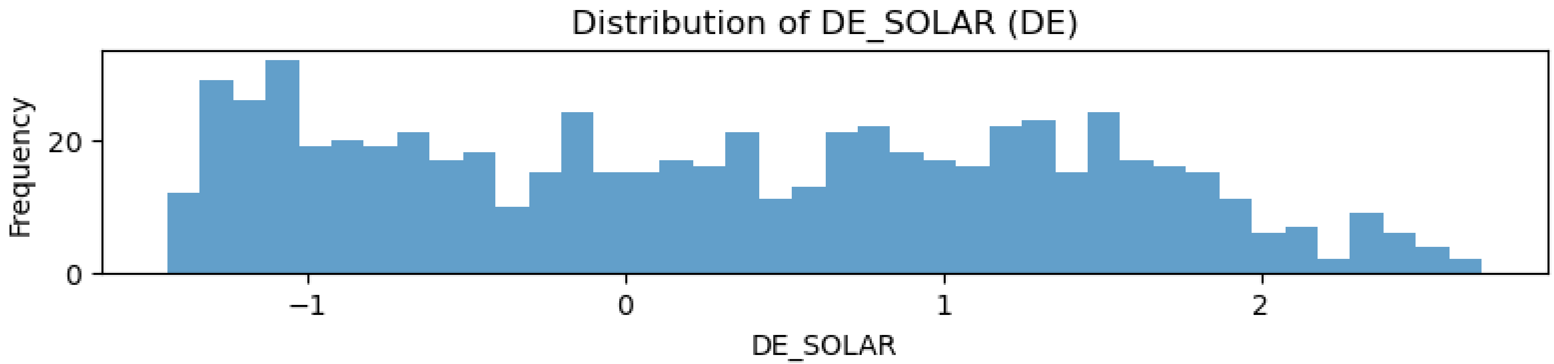
Distribution of FR_COAL (DE)
Frequency
0
200
400
−0.5
0.0
0.5
1.0
FR_COAL

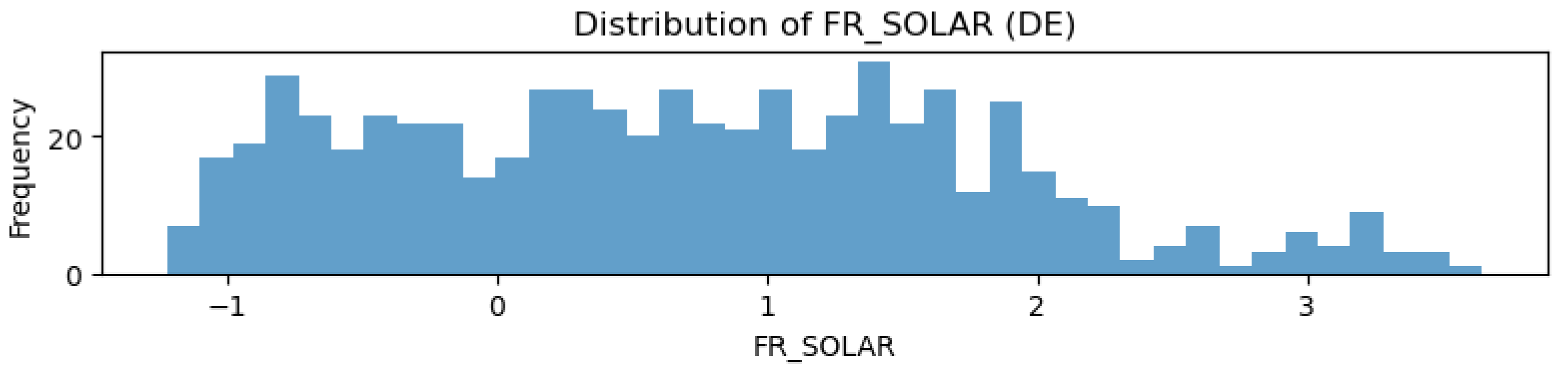

Distribution of DE_HYDRO (DE)
Frequency
0
20
40
−1
0
1
2
3
4
DE_HYDRO

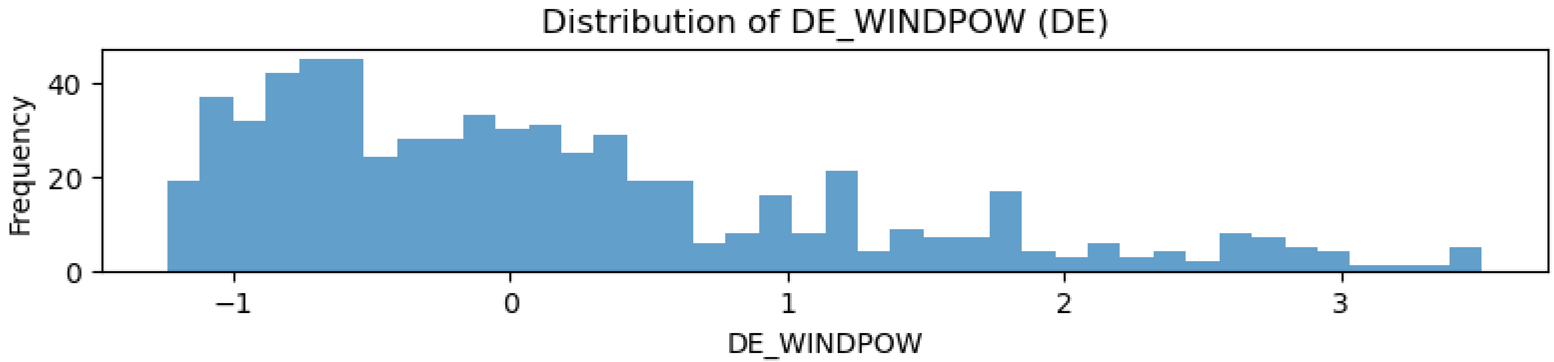

Distribution of FR_HYDRO (DE)
Frequency
0
20
40
−2
−1
0
1
2
3
FR_HYDRO

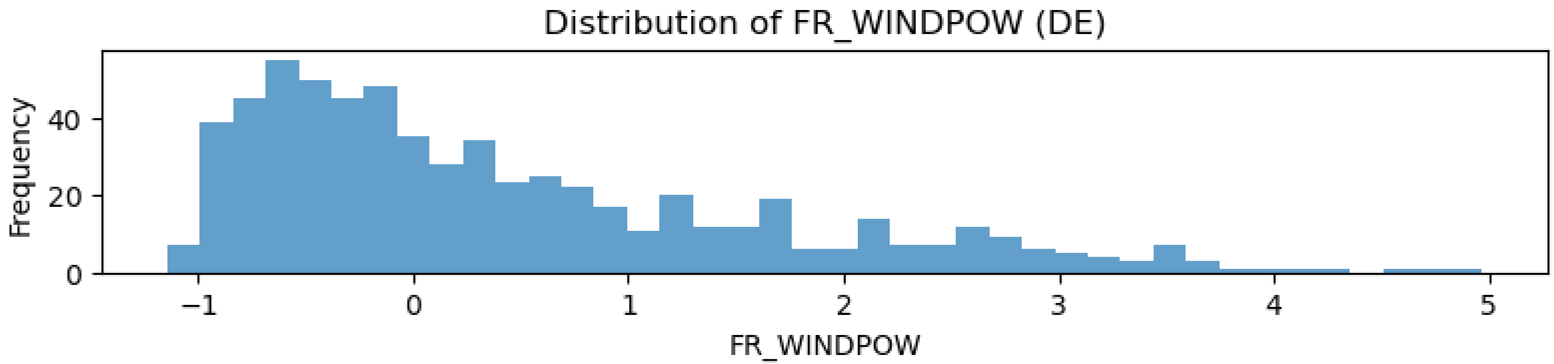

Distribution of DE_NUCLEAR (DE)
Frequency
0
50
100
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
DE_NUCLEAR

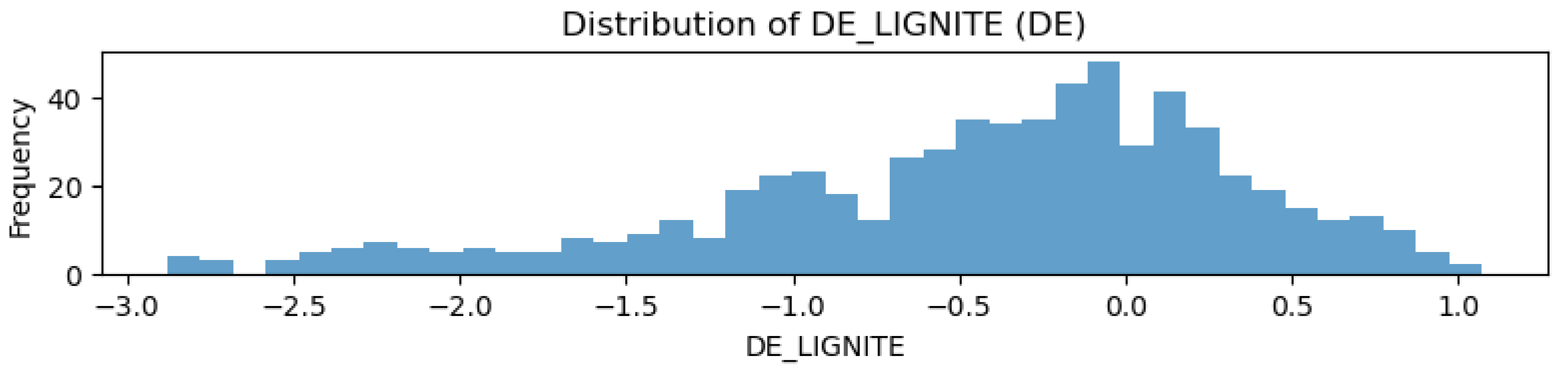

Distribution of FR_NUCLEAR (DE)
Frequency
0
20
40
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
FR_NUCLEAR

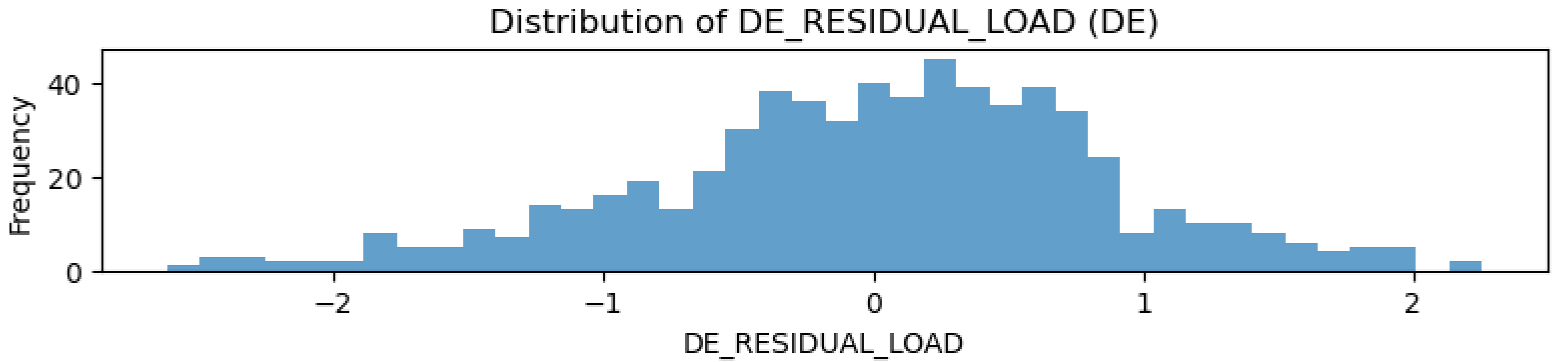

Distribution of DE_SOLAR (DE)
Frequency
0
20
−1
0
1
2
DE_SOLAR

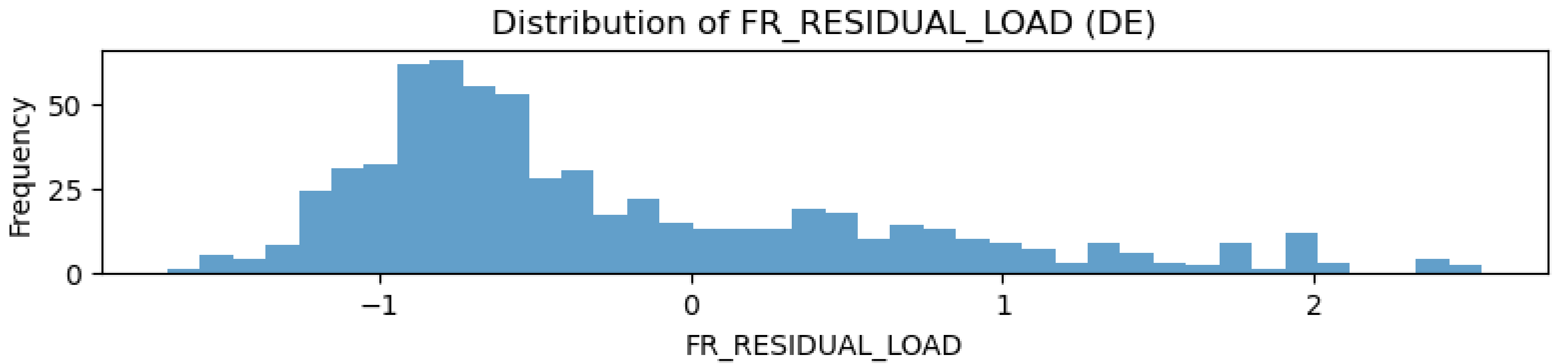

Distribution of FR_SOLAR (DE)
Frequency
0
20
−1
0
1
2
3
FR_SOLAR

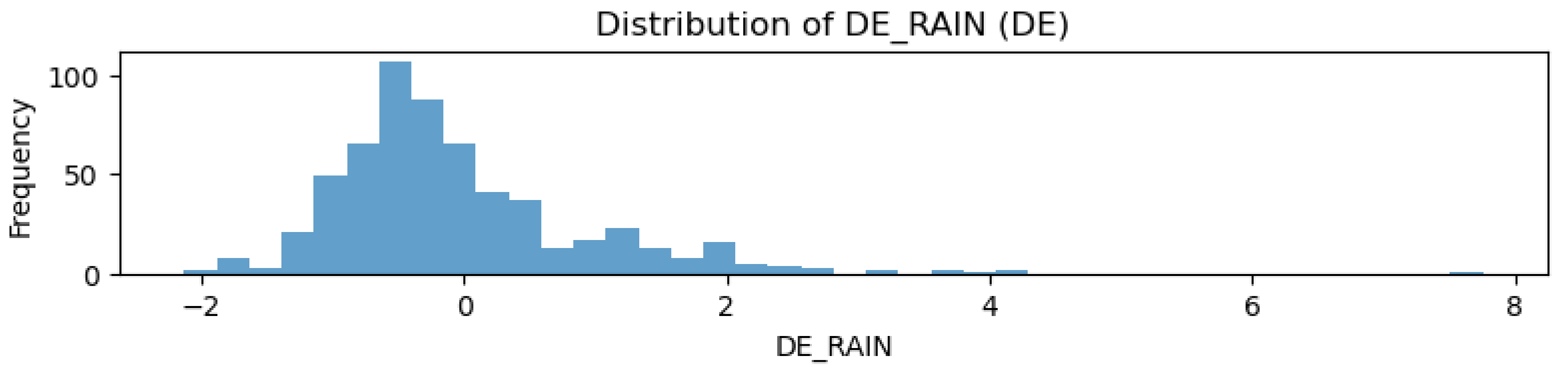

Distribution of DE_WINDPOW (DE)
Frequency
DE_WINDPOW

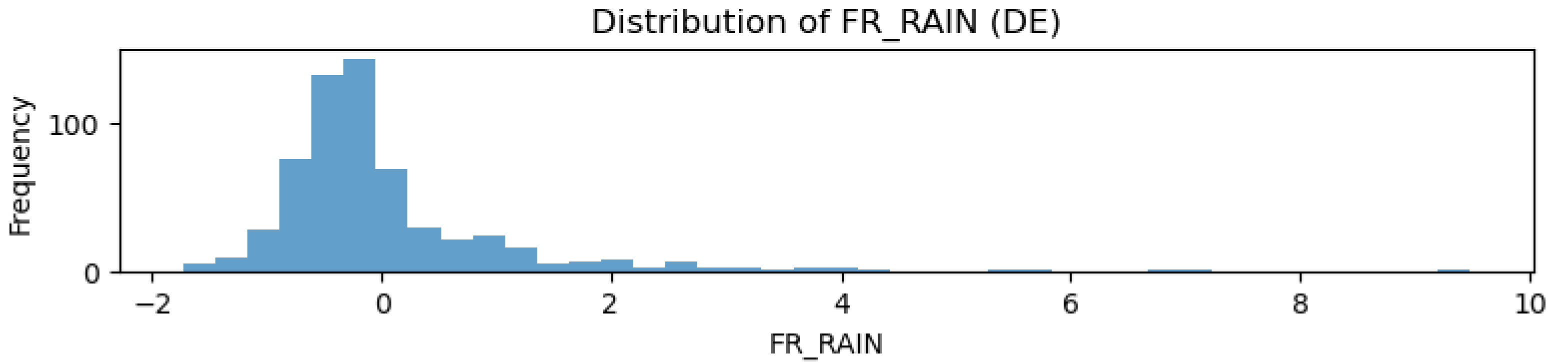

Distribution of FR_WINDPOW (DE)
Frequency
FR_WINDPOW

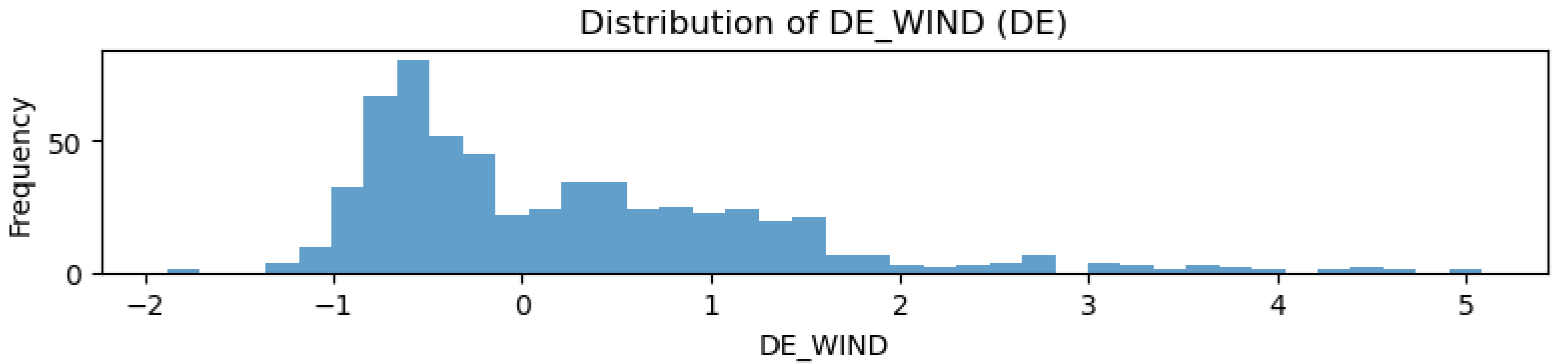

Distribution of DE_LIGNITE (DE)
Frequency
DE_LIGNITE

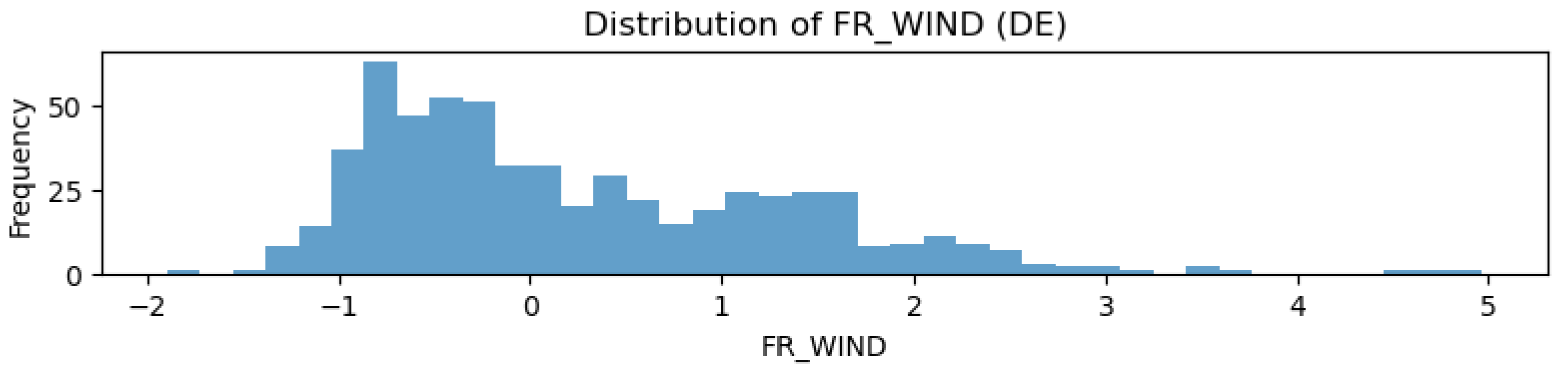

Distribution of DE_RESIDUAL_LOAD (DE)
Frequency
0
20
40
−2
−1
0
1
2
DE_RESIDUAL_LOAD

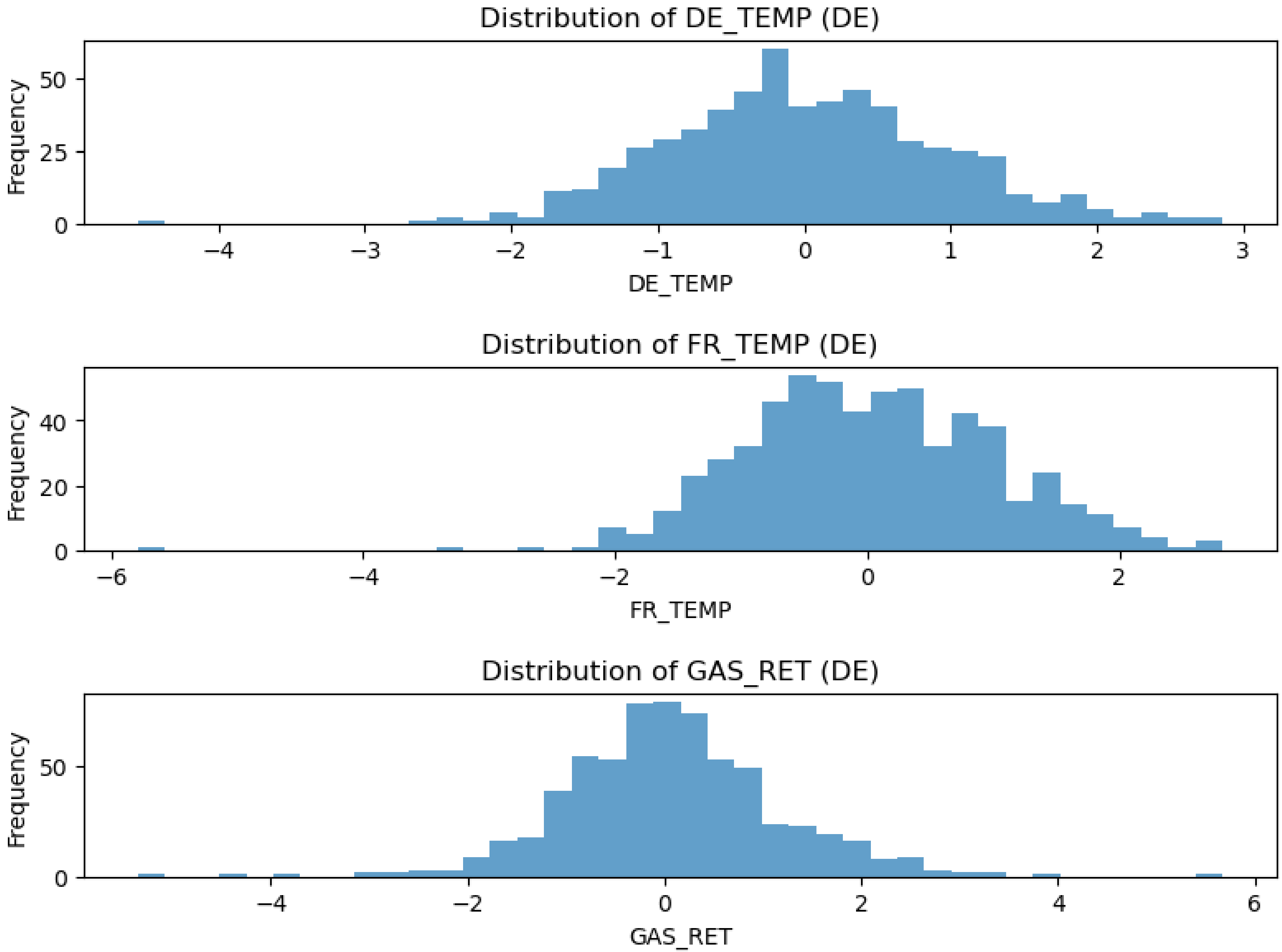

Distribution of FR_RESIDUAL_LOAD (DE)
Frequency
0
25
50
−1
0
1
2
FR_RESIDUAL_LOAD

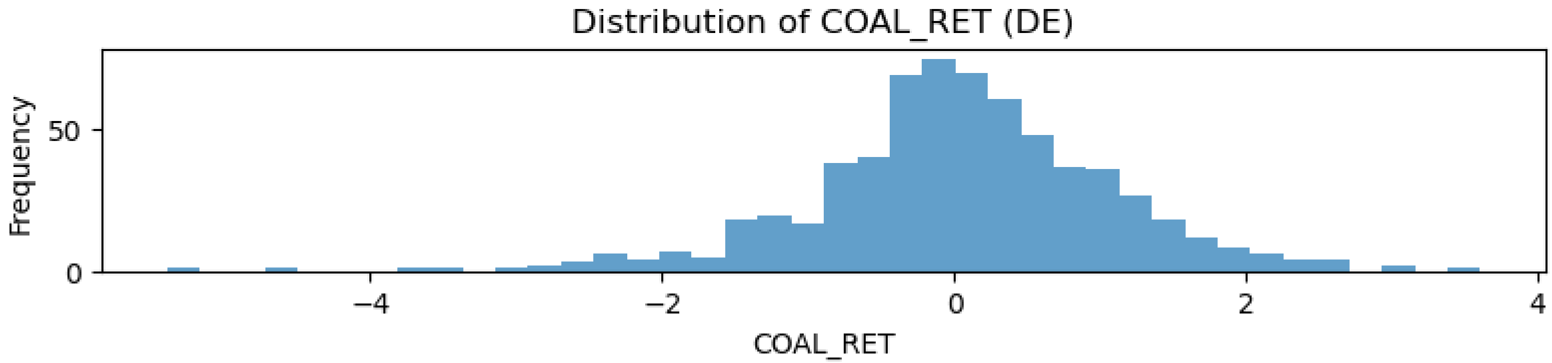

Distribution of DE_RAIN (DE)
Frequency
0
50
100
−2
0
2
4
6
8
DE_RAIN

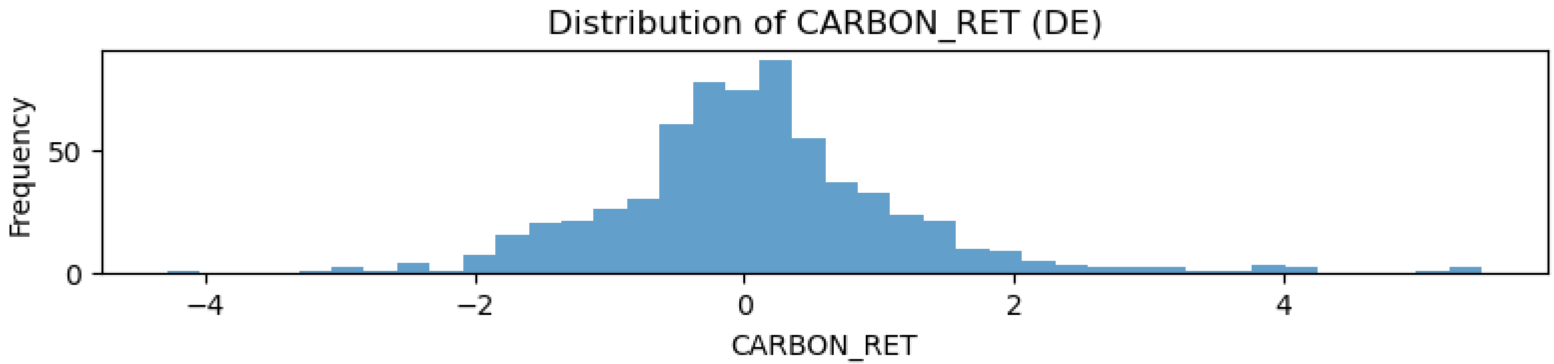

Distribution of FR_RAIN (DE)
Frequency
100
0
−2
0
2
4
6
8
10
FR_RAIN

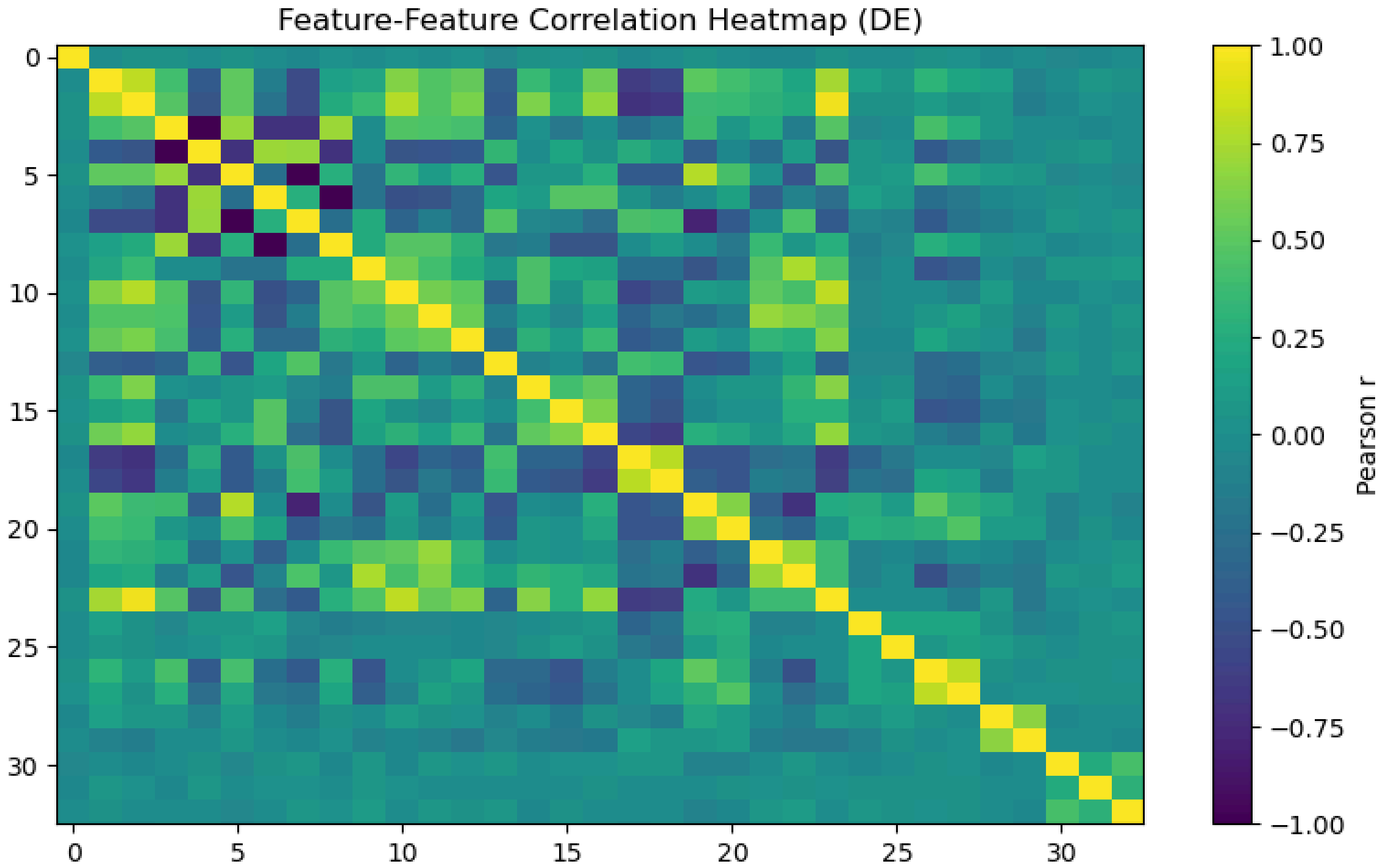

Distribution of DE_WIND (DE)
Frequency
50
0
−2
−1
0
1
2
3
4
5
DE_WIND


Distribution of FR_WIND (DE)
Frequency
50
25
0
−2
−1
0
1
2
3
4
5
FR_WIND

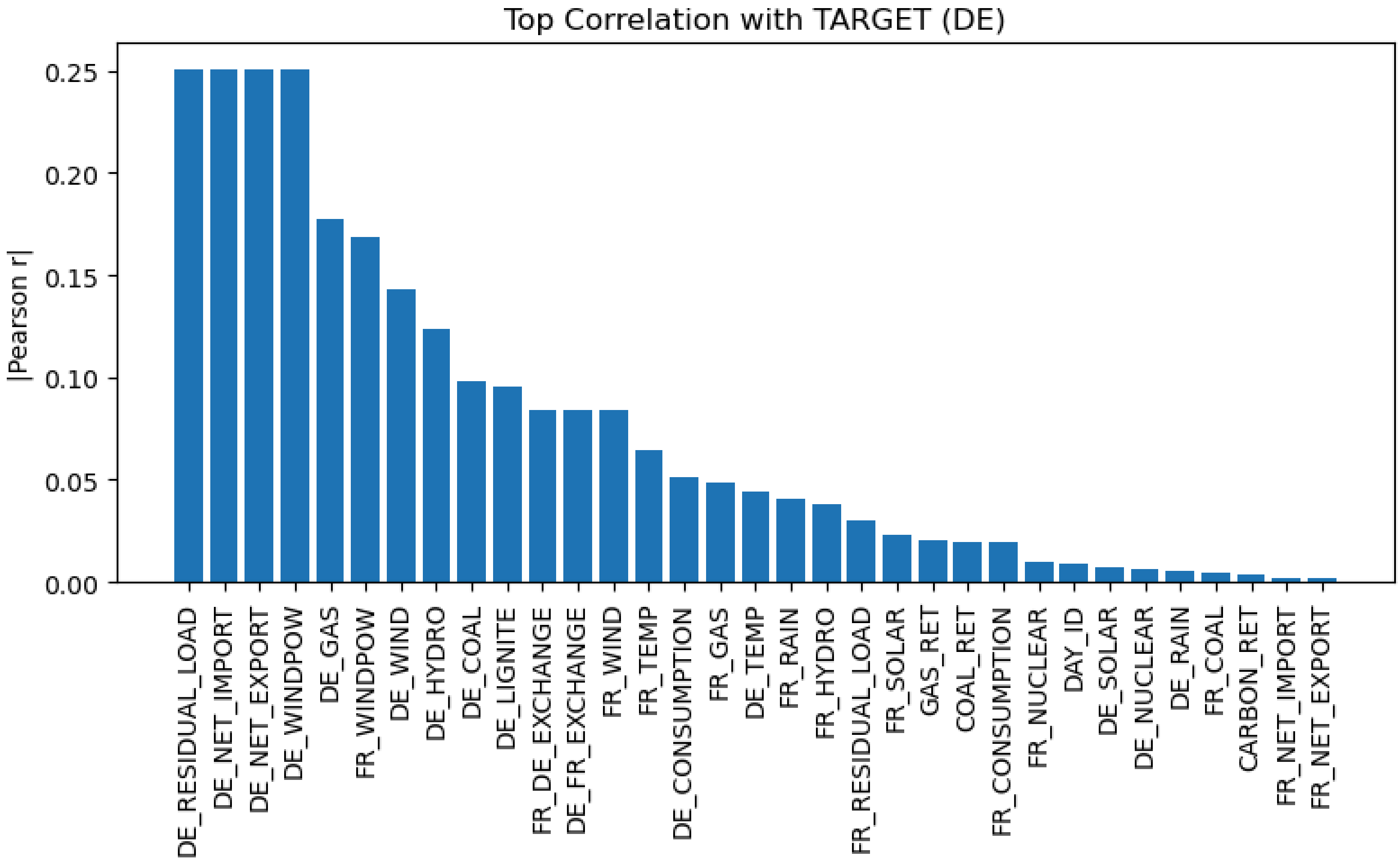

Distribution of DE_TEMP (DE)
Frequency
50
25
0
−4
−3
−2
−1
0
1
2
3
DE_TEMP
Distribution of FR_TEMP (DE)
Frequency
40
20
0
−6
−4
−2
0
2
FR_TEMP
Distribution of GAS_RET (DE)
Frequency
50
0
−4
−2
0
2
4
6
GAS_RET


Distribution of COAL_RET (DE)
Frequency
50
0
−4
−2
0
2
4
COAL_RET


Distribution of CARBON_RET (DE)
Frequency
50
0
−4
−2
0
2
4
CARBON_RET


Feature-Feature Correlation Heatmap (DE)
0
5
10
15
20
25
30
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Pearson r

```
DE_RESIDUAL_LOAD    0.250898
DE_NET_IMPORT       0.250616
DE_GAS              0.177561
DE_HYDRO            0.123676
DE_COAL             0.097803
DE_LIGNITE          0.095619
FR_DE_EXCHANGE      0.084194
FR_GAS              0.048483
FR_HYDRO            0.038172
FR_RESIDUAL_LOAD    0.029686
FR_SOLAR            0.022717
GAS_RET             0.020087
FR_NUCLEAR          0.009855
DAY_ID              0.008573
DE_SOLAR            0.007162
FR_COAL             0.004934
CARBON_RET          0.003538
FR_NET_EXPORT       0.001544
FR_NET_IMPORT      -0.001544
DE_RAIN            -0.005381
DE_NUCLEAR         -0.006729
FR_CONSUMPTION     -0.019582
COAL_RET           -0.019866
FR_RAIN            -0.040447
DE_TEMP            -0.044221
DE_CONSUMPTION     -0.051182
FR_TEMP            -0.064272
FR_WIND            -0.083825
DE_FR_EXCHANGE     -0.084194
DE_WIND            -0.143149
FR_WINDPOW         -0.168554
DE_WINDPOW         -0.250204
DE_NET_EXPORT      -0.250616
Name: TARGET, dtype: float64
```



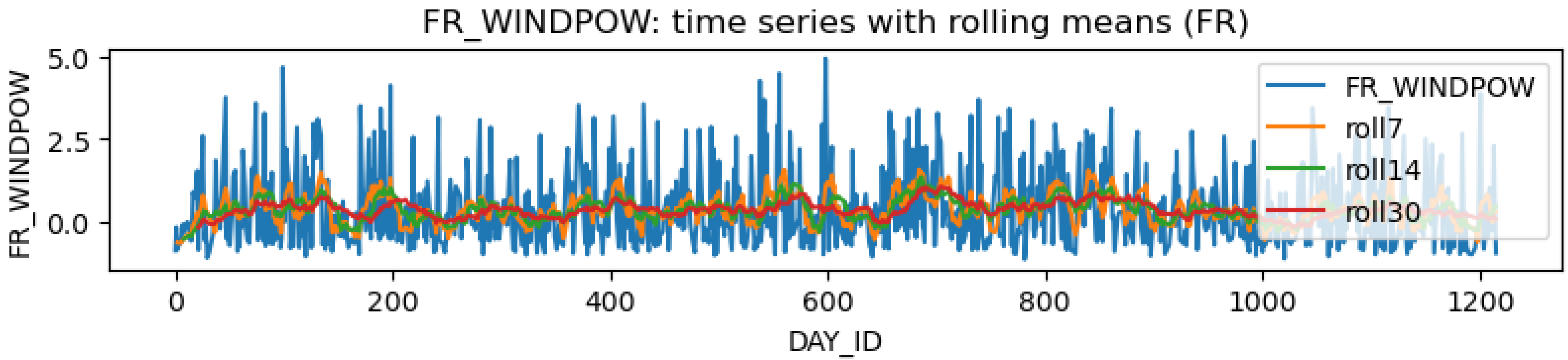
9. Time Series per Country (FR / DR)

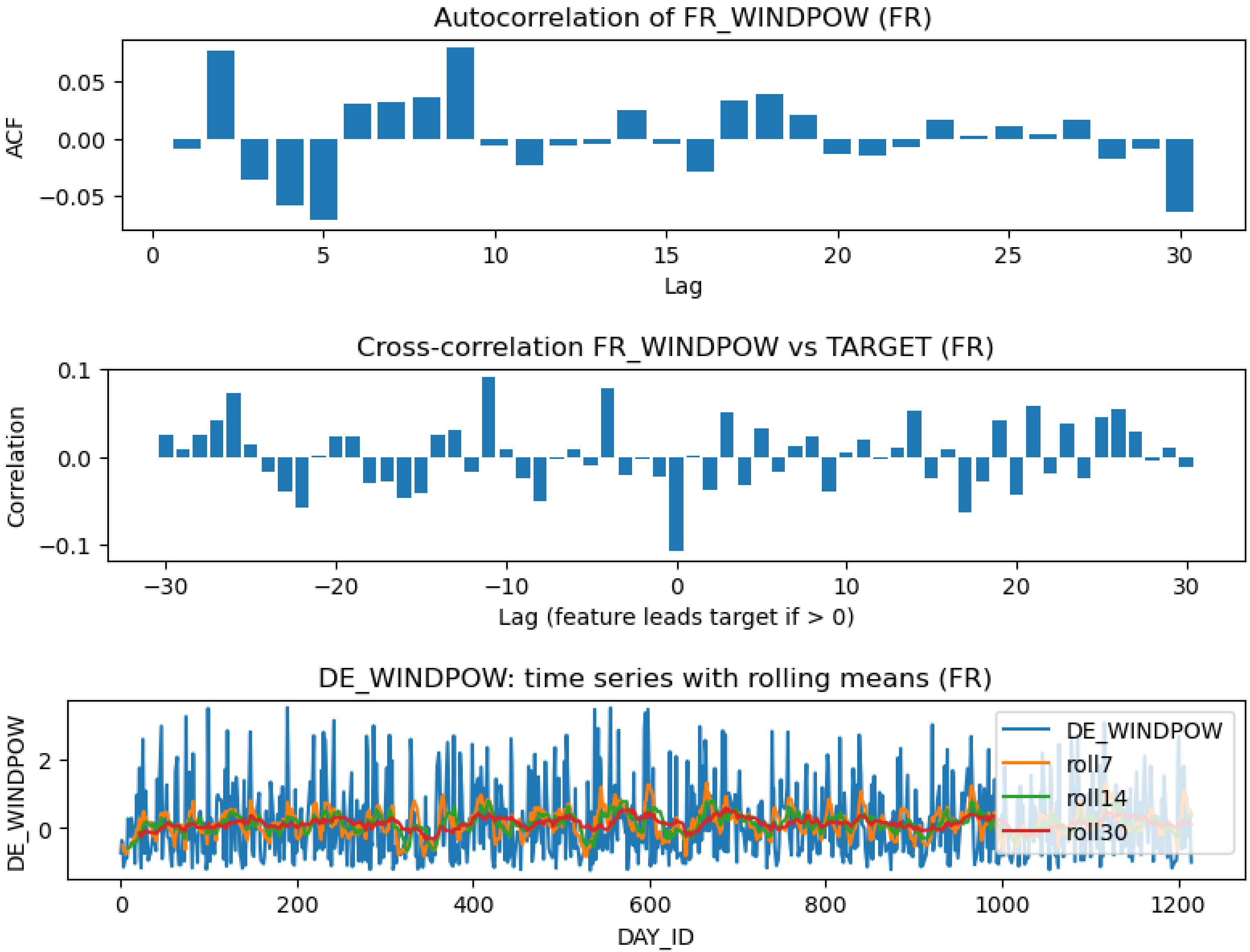
In [18]:
```
if "COUNTRY" not in df.columns:
    print("COUNTRY not found; skipping per-country TS analysis.")
else:
    for country in df["COUNTRY"].dropna().unique():
        print("\n==============================")
        print(f"### COUNTRY: {country}")
        d = df[df["COUNTRY"]==country].copy()
```

```
    if "DAY_ID" not in d.columns:
        print("DAY_ID not available; skip TS plots.")
        continue
    d = d.sort_values("DAY_ID")

    # Select features by |corr| within this country
    numc = [c for c in d.columns if c not in {"ID","DAY_ID","COUNTRY","TARGET"} and pd.api.types.is_numeric_dtype(d[c])]
    if "TARGET" not in d.columns or len(numc)==0:
        print("Skip (no target or no numeric cols).")
        continue
    corr_to_t = d[numc + ["TARGET"]].corr()["TARGET"].drop("TARGET").abs().sort_values(ascending=False)
    ts_feats_c = list(corr_to_t.index[:len(corr_to_t)])
    print("Selected features:", ts_feats_c)
    yv = d["TARGET"].astype(float).to_numpy()

    for c in ts_feats_c:
        # TS + rolling
        s = d[c].astype(float)
        plt.figure()
        plt.subplot(311)
        plt.plot(d["DAY_ID"], s, label=f"{c}")
        for w in ROLL_WINDOWS:
            try:
                plt.plot(d["DAY_ID"], s.rolling(window=w, min_periods=max(1,w//2)).mean(), label=f"roll{w}")
            except Exception:
                pass
        plt.xlabel("DAY_ID"); plt.ylabel(c)
        plt.title(f"{c}: time series with rolling means ({country})")
        plt.legend()
        plt.tight_layout()
        plt.show()

        # ACF
        vals = d[c].astype(float).to_numpy()
        acf = acf_values(vals)
        plt.subplot(312)
        plt.bar(np.arange(1, len(acf)+1), acf)
        plt.xlabel("Lag"); plt.ylabel("ACF")
        plt.title(f"Autocorrelation of {c} ({country})")
        plt.tight_layout()
        plt.show()
```

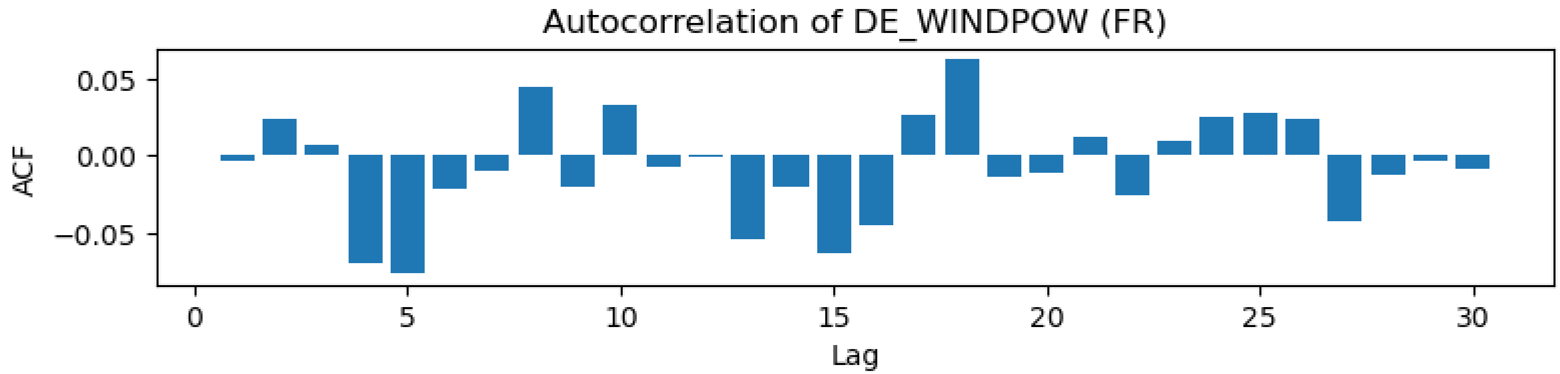
```
        # Cross-corr with TARGET
        xv = d[c].astype(float).to_numpy()
        lags, cc = lagged_corr(xv, yv)
        plt.subplot(313)
        plt.bar(lags, cc)
        plt.xlabel("Lag (feature leads target if > 0)")
        plt.ylabel("Correlation")
        plt.title(f"Cross-correlation {c} vs TARGET ({country})")
        plt.tight_layout()
        plt.show()
```

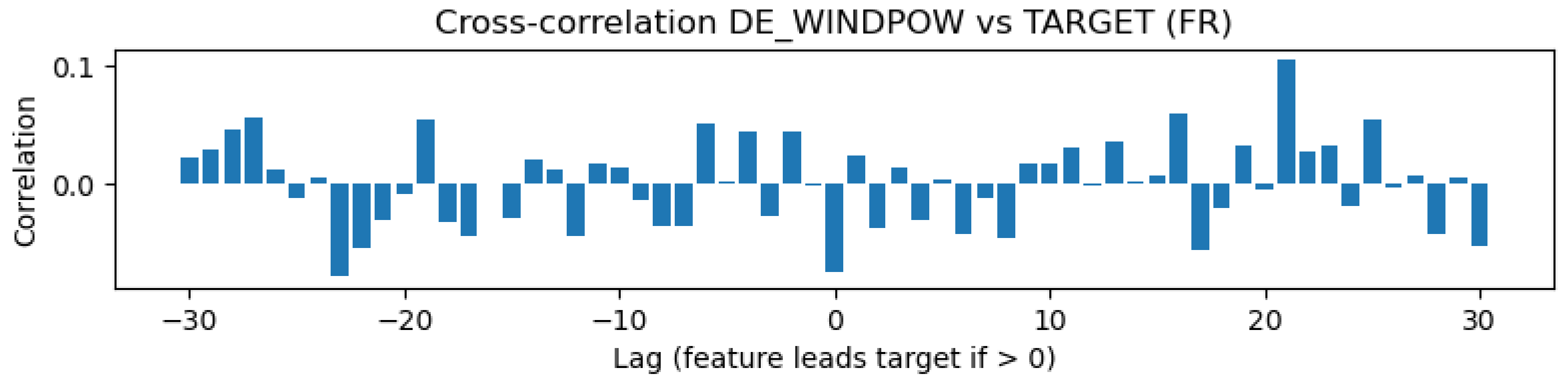
```
==============================
### COUNTRY: FR
Selected features: ['FR_WINDPOW', 'DE_WINDPOW', 'FR_HYDRO', 'DE_NET_IMPORT', 'DE_NET_EXPORT', 'DE_RAIN', 'CARBON_RET', 'GAS_RE
T', 'DE_HYDRO', 'DE_CONSUMPTION', 'DE_RESIDUAL_LOAD', 'DE_WIND', 'DE_TEMP', 'FR_NET_EXPORT', 'FR_NET_IMPORT', 'FR_RAIN', 'FR_RE
SIDUAL_LOAD', 'DE_NUCLEAR', 'FR_DE_EXCHANGE', 'DE_FR_EXCHANGE', 'FR_TEMP', 'FR_WIND', 'DE_GAS', 'FR_COAL', 'DE_SOLAR', 'DE_COA
L', 'FR_SOLAR', 'FR_CONSUMPTION', 'FR_GAS', 'FR_NUCLEAR', 'DE_LIGNITE', 'COAL_RET']
```

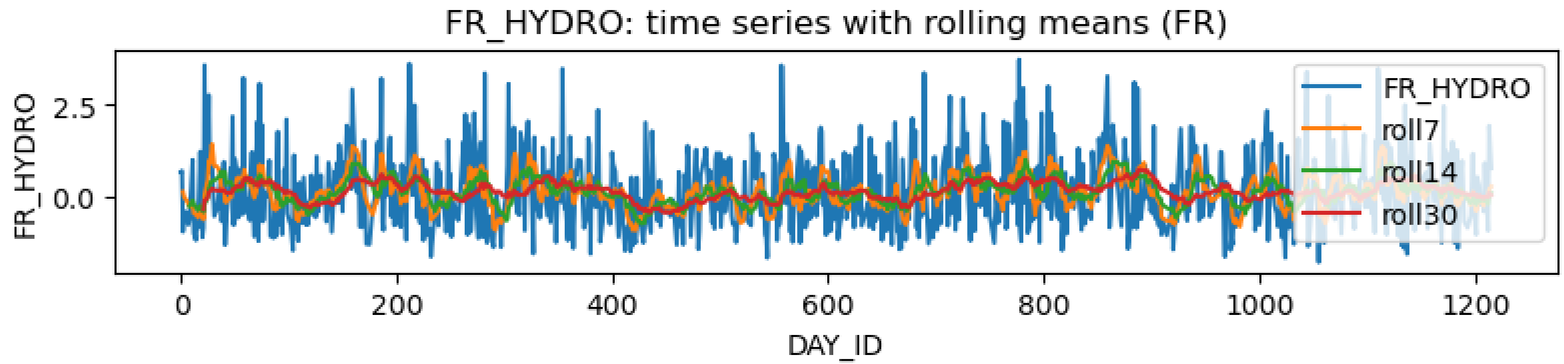


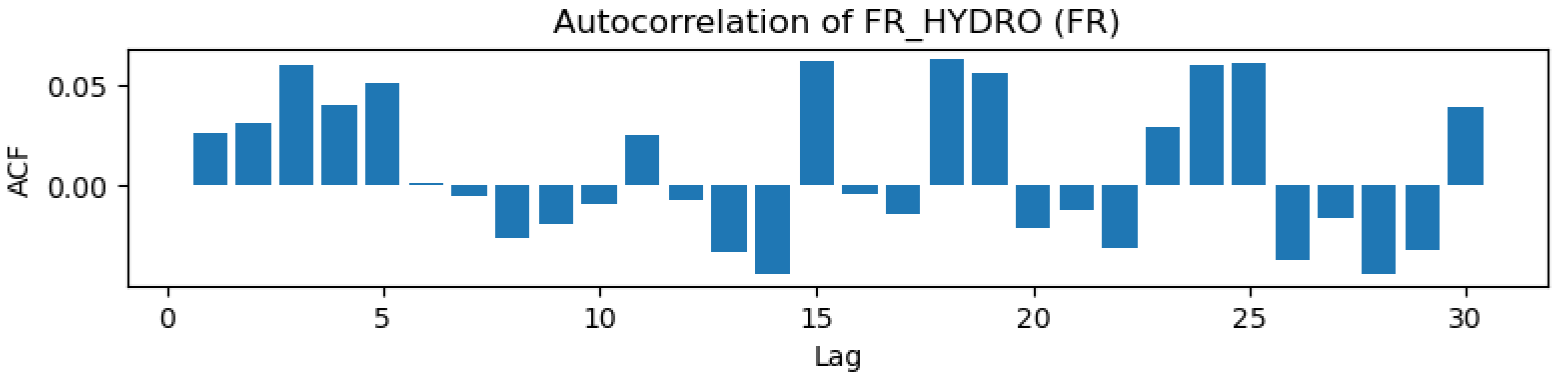


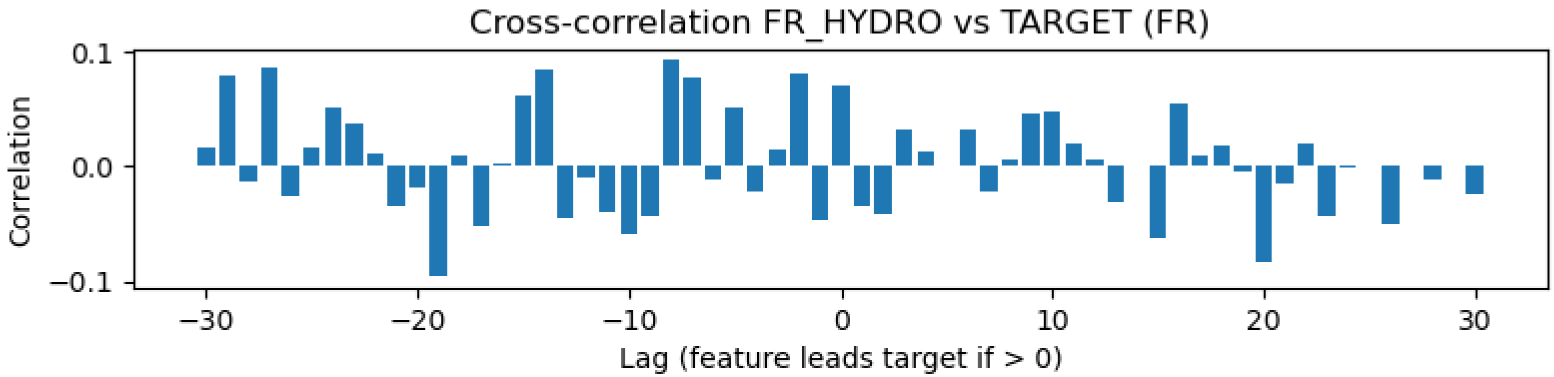
Autocorrelation of FR_WINDPOW (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag
Cross-correlation FR_WINDPOW vs TARGET (FR)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)
DE_WINDPOW: time series with rolling means (FR)
DE_WINDPOW
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_WINDPOW
roll7
roll14
roll30

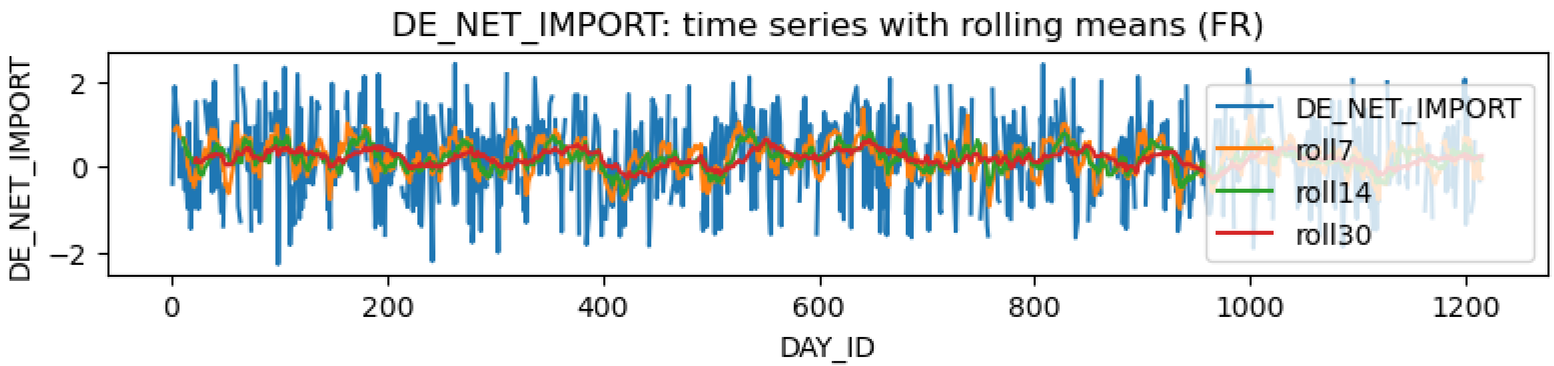

Autocorrelation of DE_WINDPOW (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

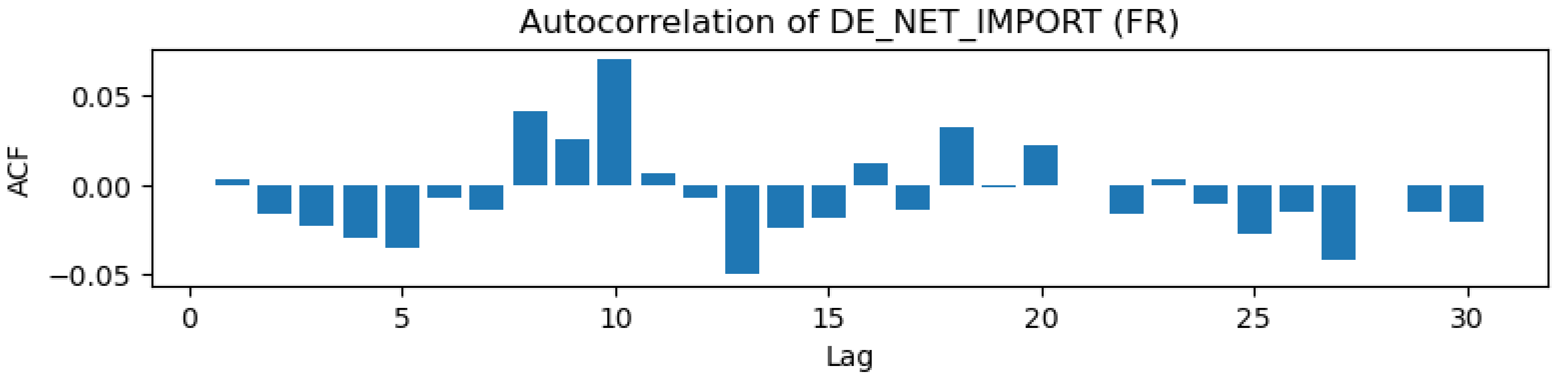

Cross-correlation DE_WINDPOW vs TARGET (FR)
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

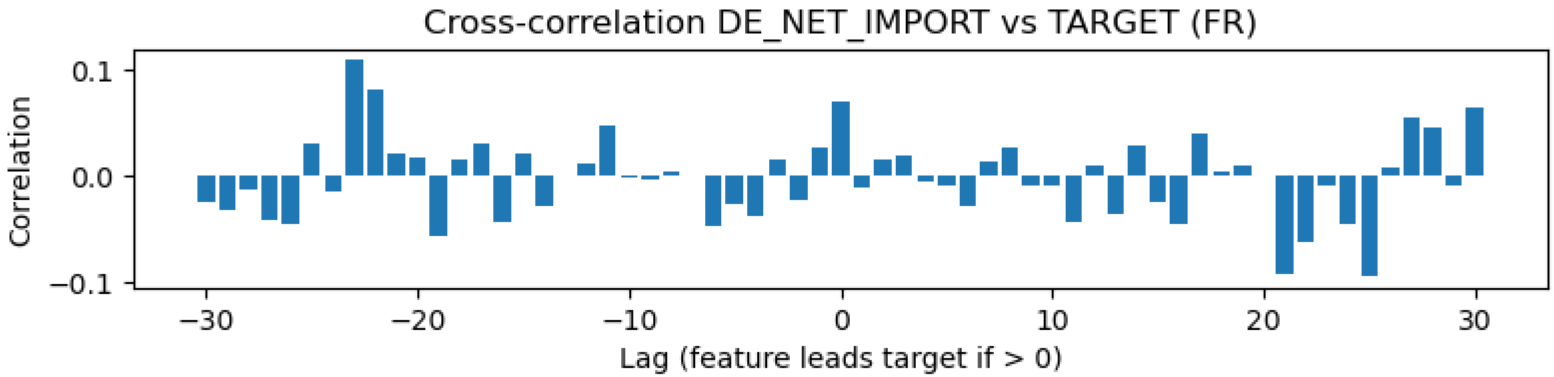

FR_HYDRO: time series with rolling means (FR)
FR_HYDRO
2.5
0.0
FR_HYDRO
roll7
roll14
roll30
0
200
400
600
800
1000
1200
DAY_ID

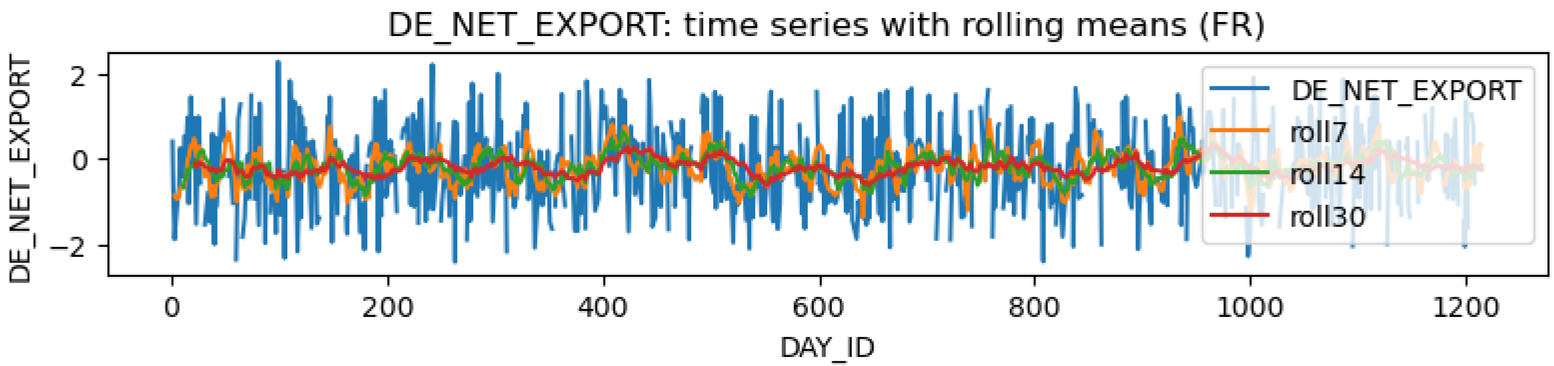

Autocorrelation of FR_HYDRO (FR)
ACF
0.05
0.00
0
5
10
15
20
25
30
Lag

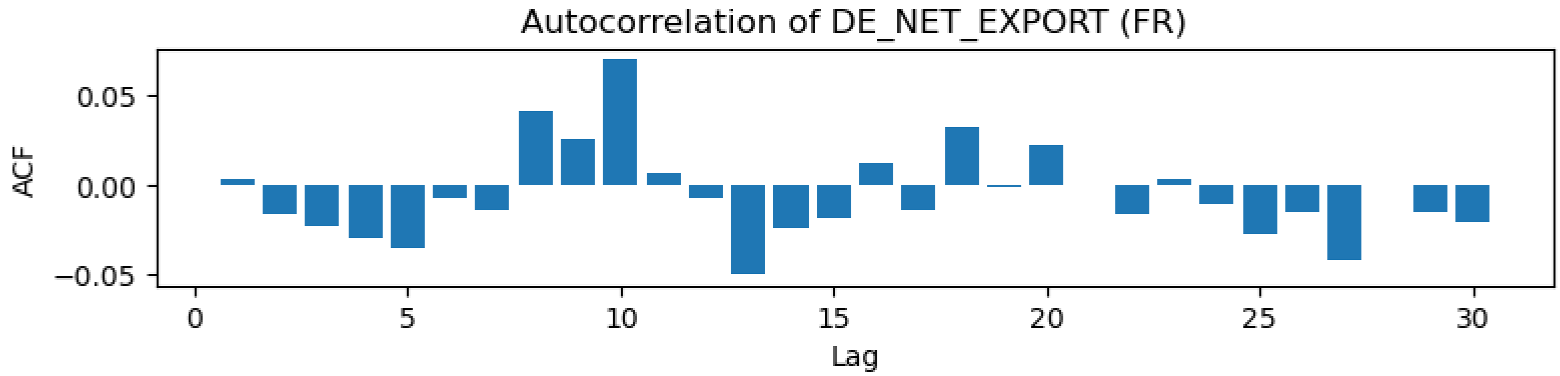

Cross-correlation FR_HYDRO vs TARGET (FR)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

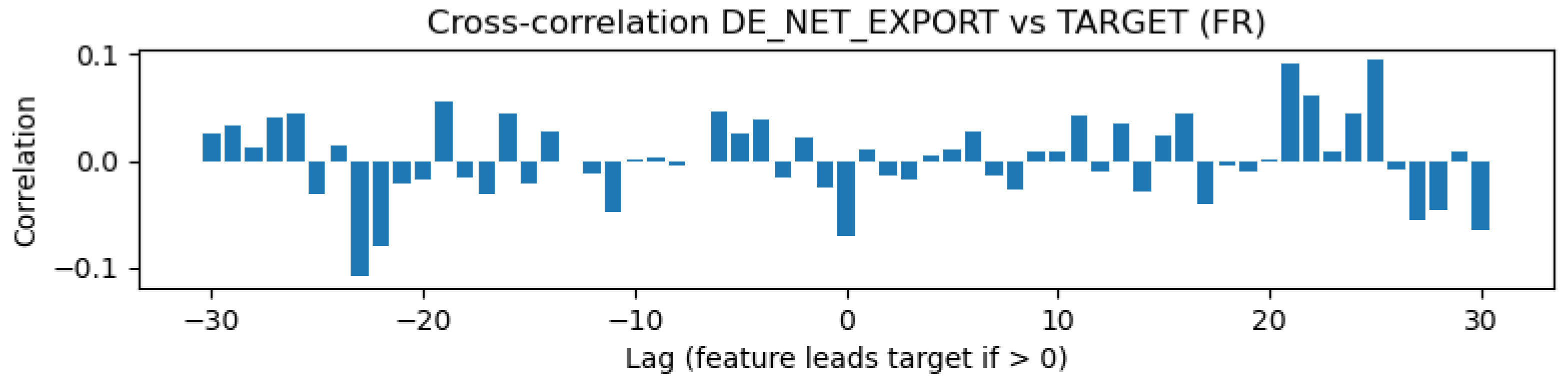

DE_NET_IMPORT: time series with rolling means (FR)
DE_NET_IMPORT
2
0
−2
DE_NET_IMPORT
roll7
roll14
roll30
0
200
400
600
800
1000
1200
DAY_ID

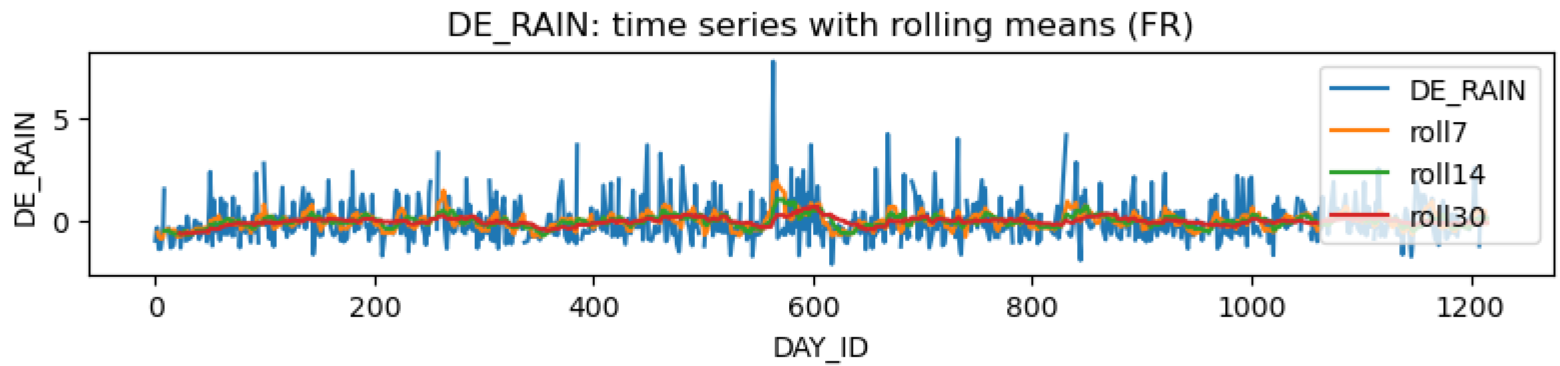

Autocorrelation of DE_NET_IMPORT (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

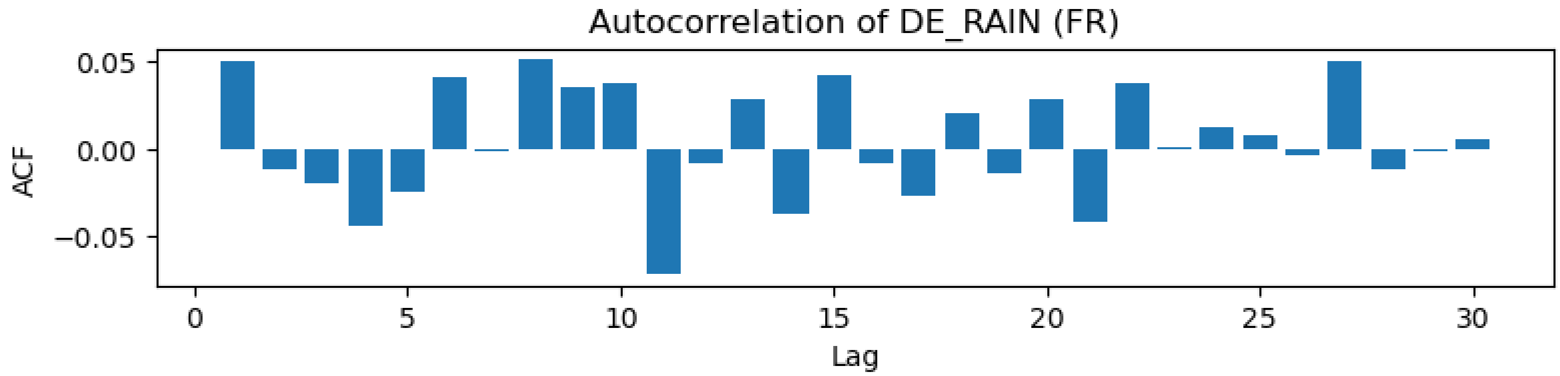

Cross-correlation DE_NET_IMPORT vs TARGET (FR)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

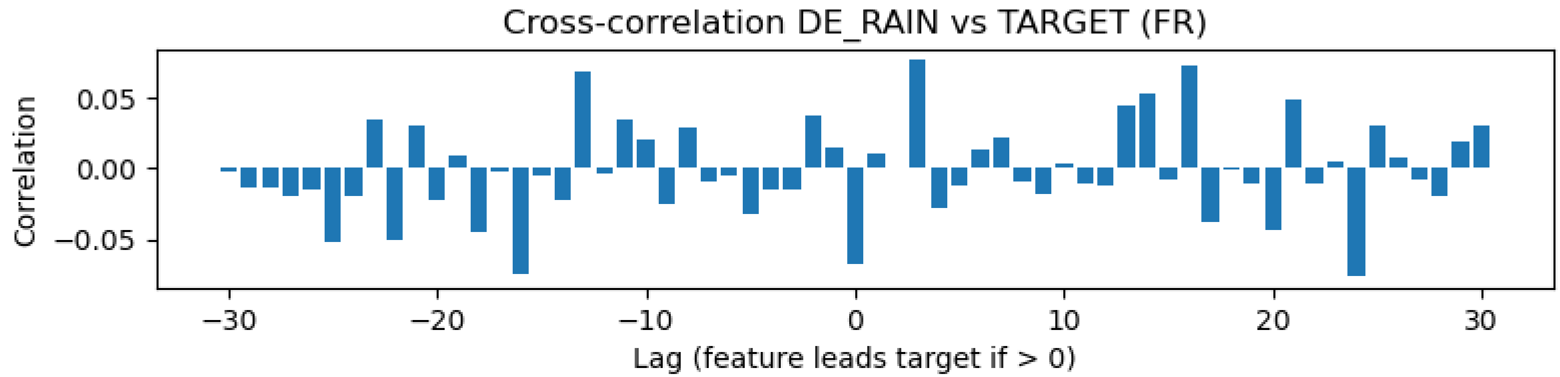

DE_NET_EXPORT: time series with rolling means (FR)
DE_NET_EXPORT
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_NET_EXPORT
roll7
roll14
roll30

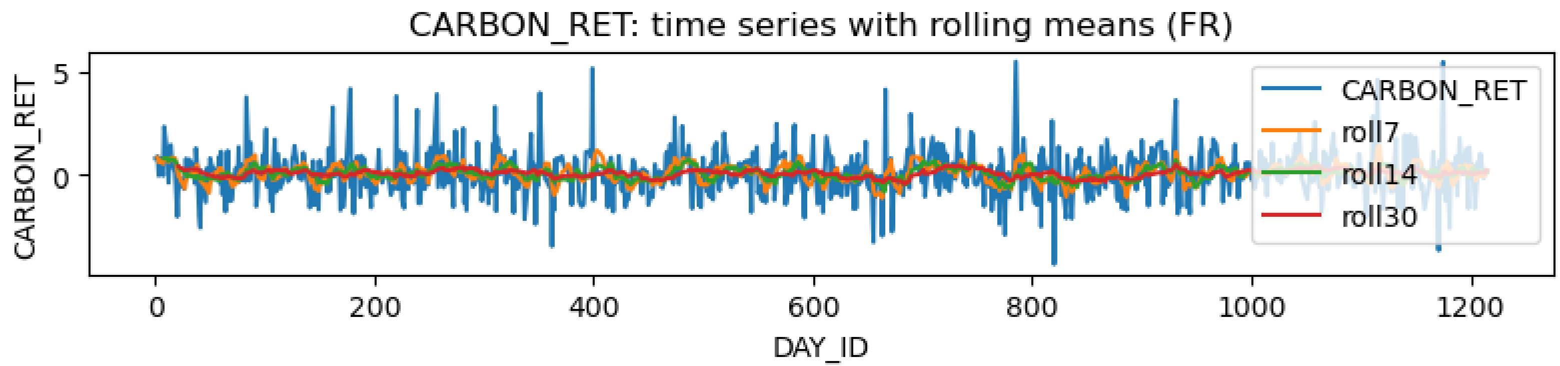

Autocorrelation of DE_NET_EXPORT (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

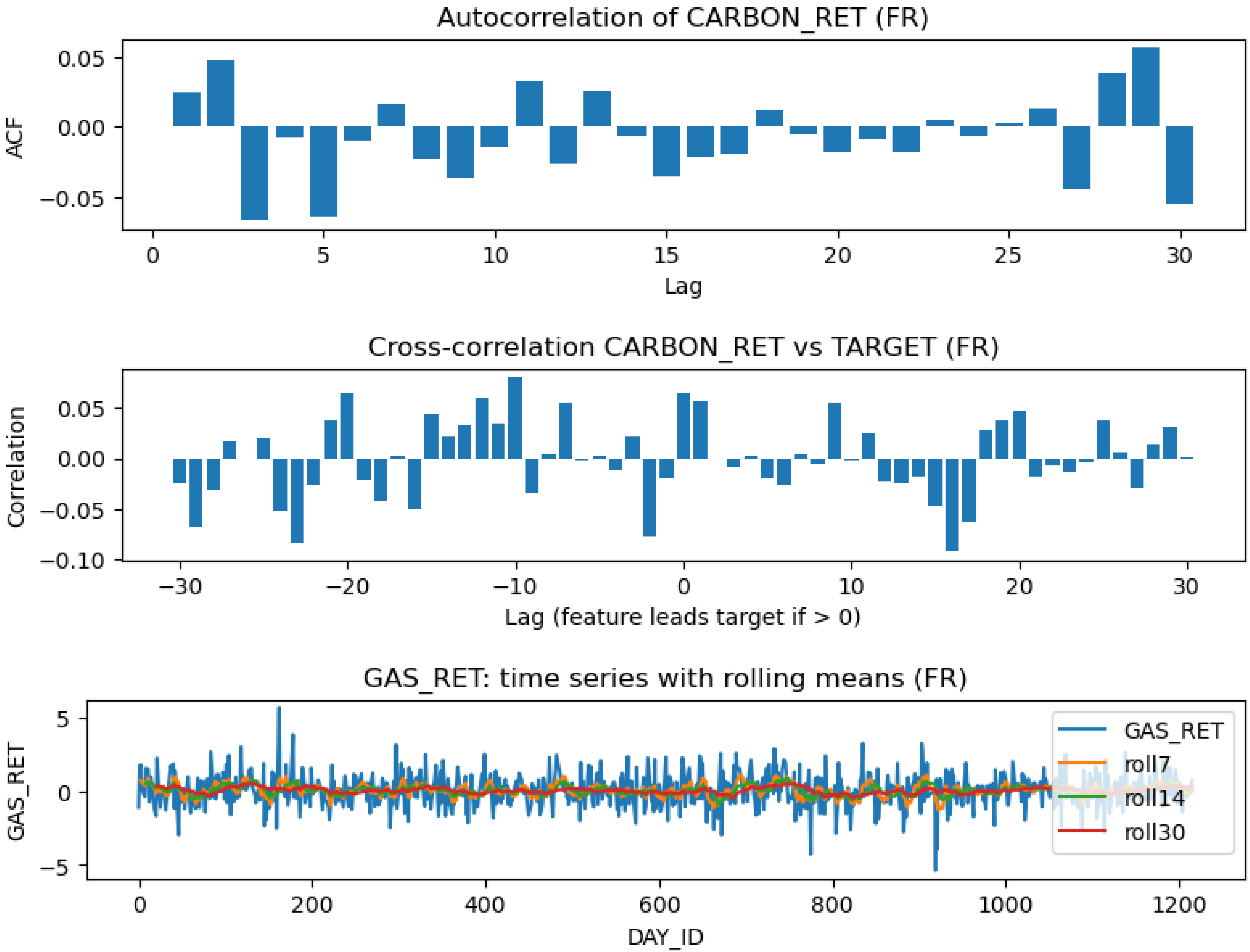

Cross-correlation DE_NET_EXPORT vs TARGET (FR)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

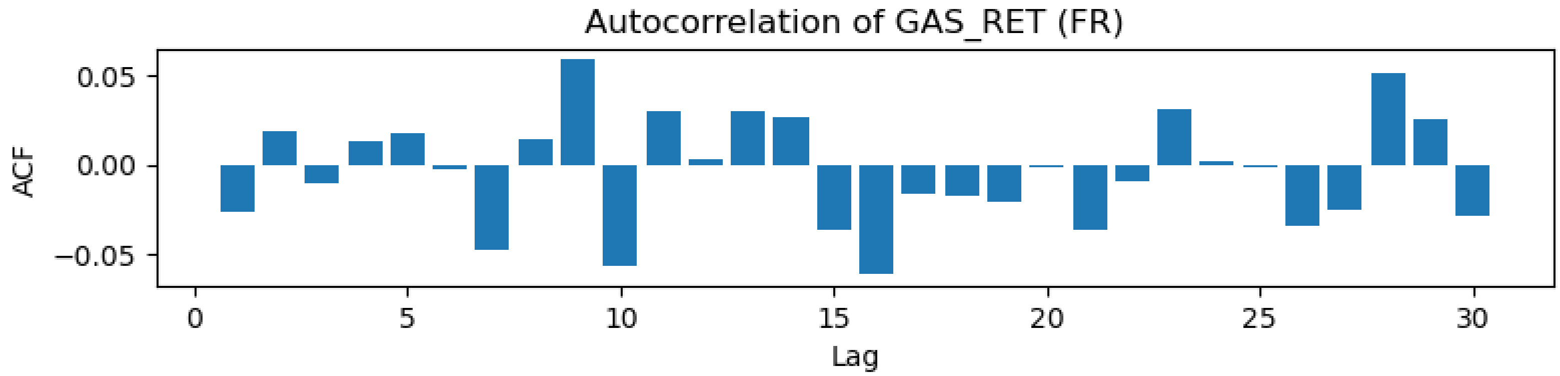

DE_RAIN: time series with rolling means (FR)
DE_RAIN
5
0
0
200
400
600
800
1000
1200
DAY_ID
DE_RAIN
roll7
roll14
roll30

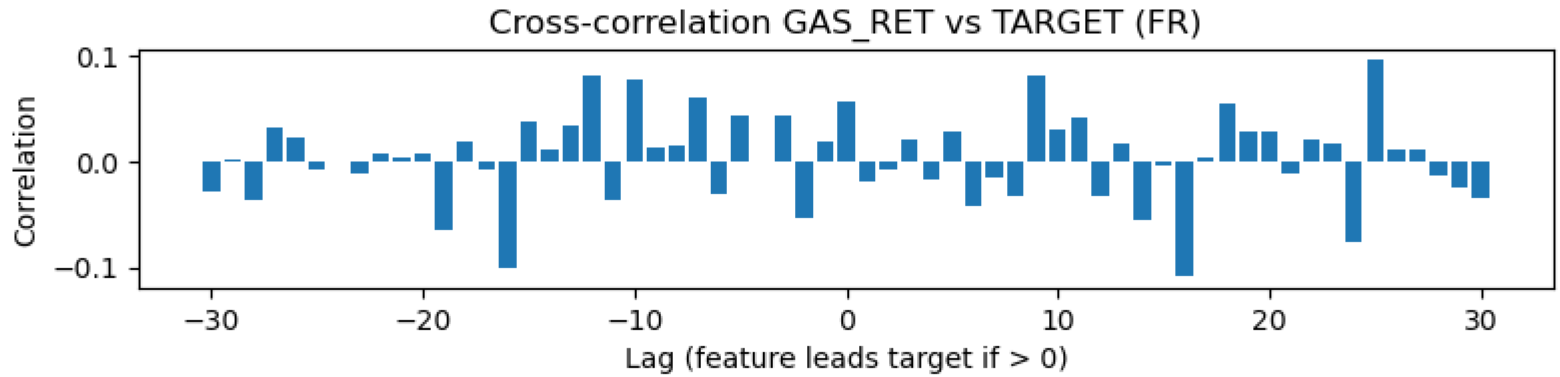

Autocorrelation of DE_RAIN (FR)
ACF
Lag
0.05
0.00
−0.05
0
5
10
15
20
25
30

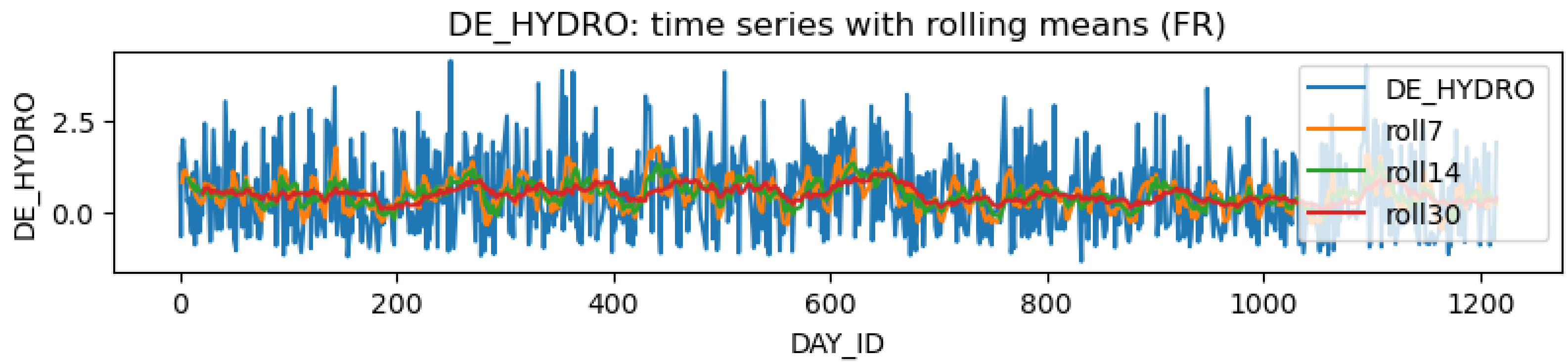

Cross-correlation DE_RAIN vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30

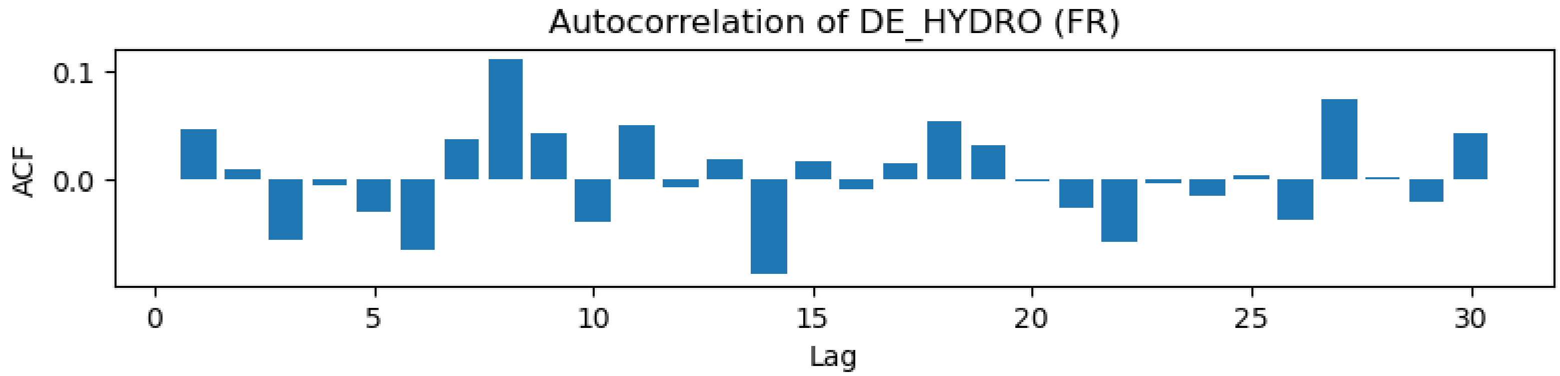

CARBON_RET: time series with rolling means (FR)
CARBON_RET
DAY_ID
5
0
0
200
400
600
800
1000
1200
CARBON_RET
roll7
roll14
roll30

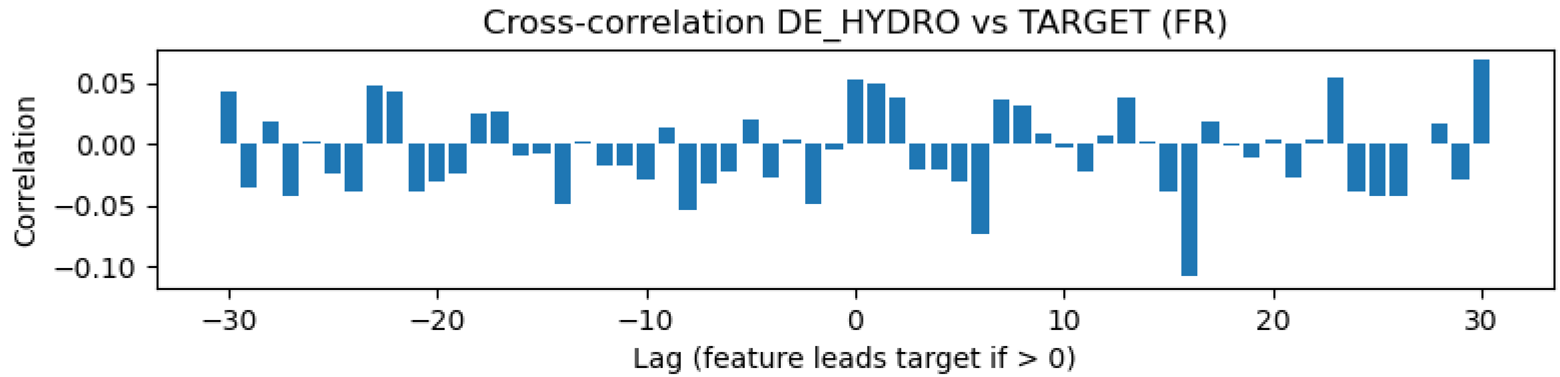

Autocorrelation of CARBON_RET (FR)
ACF
Lag
Cross-correlation CARBON_RET vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)
GAS_RET: time series with rolling means (FR)
GAS_RET
DAY_ID
GAS_RET
roll7
roll14
roll30

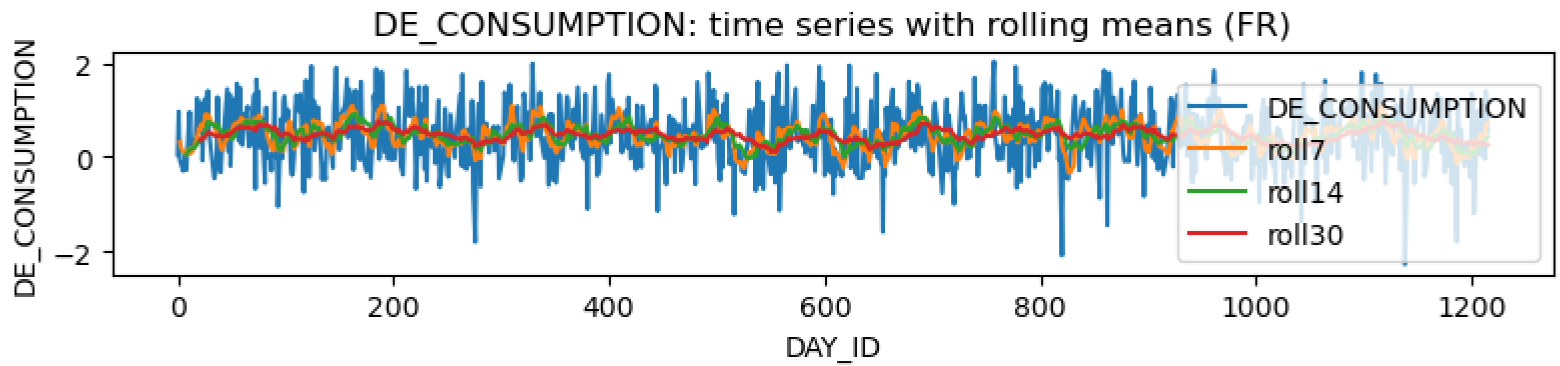

Autocorrelation of GAS_RET (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

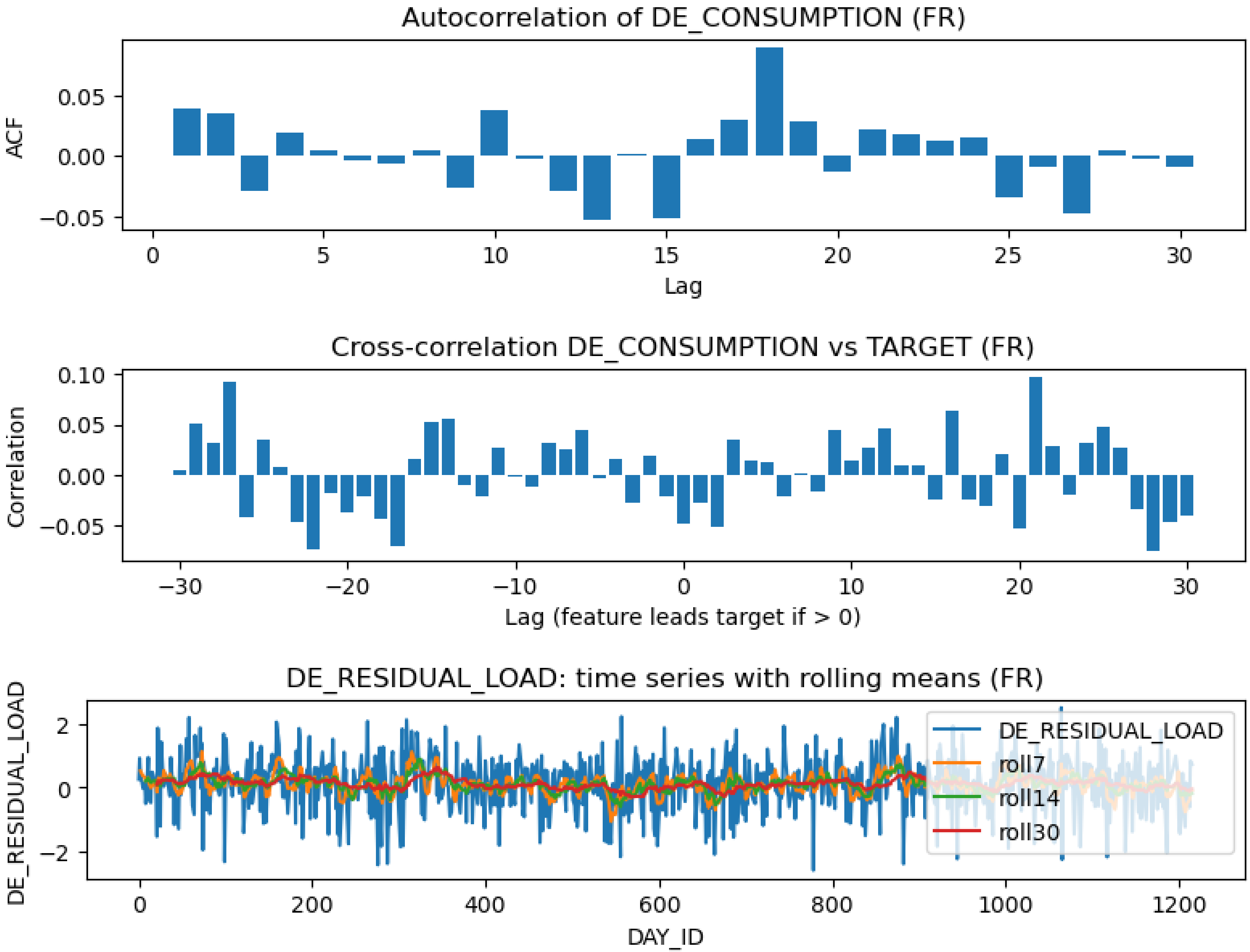

Cross-correlation GAS_RET vs TARGET (FR)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

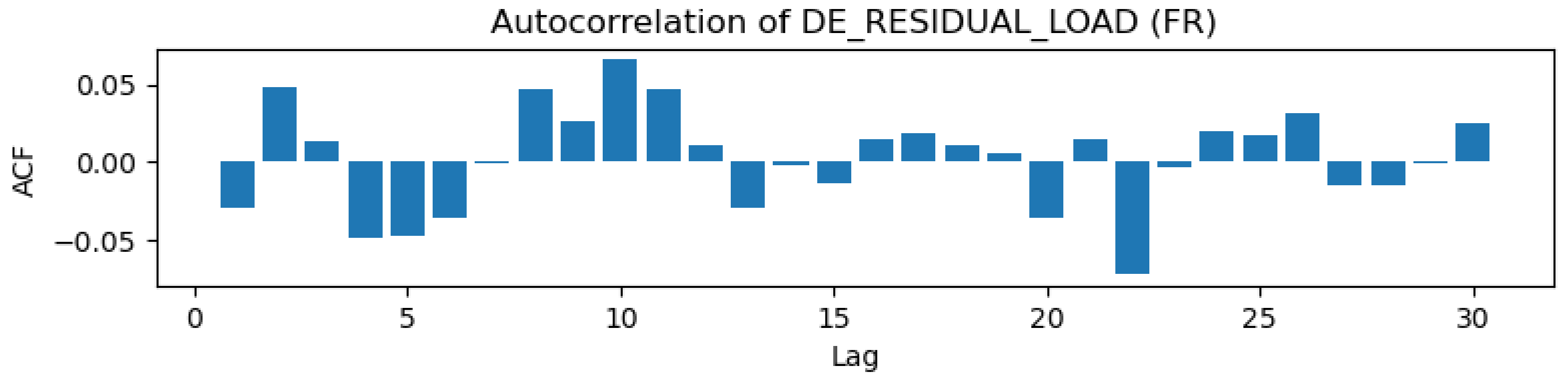

DE_HYDRO: time series with rolling means (FR)
DE_HYDRO
2.5
0.0
0
200
400
600
800
1000
1200
DAY_ID
DE_HYDRO
roll7
roll14
roll30

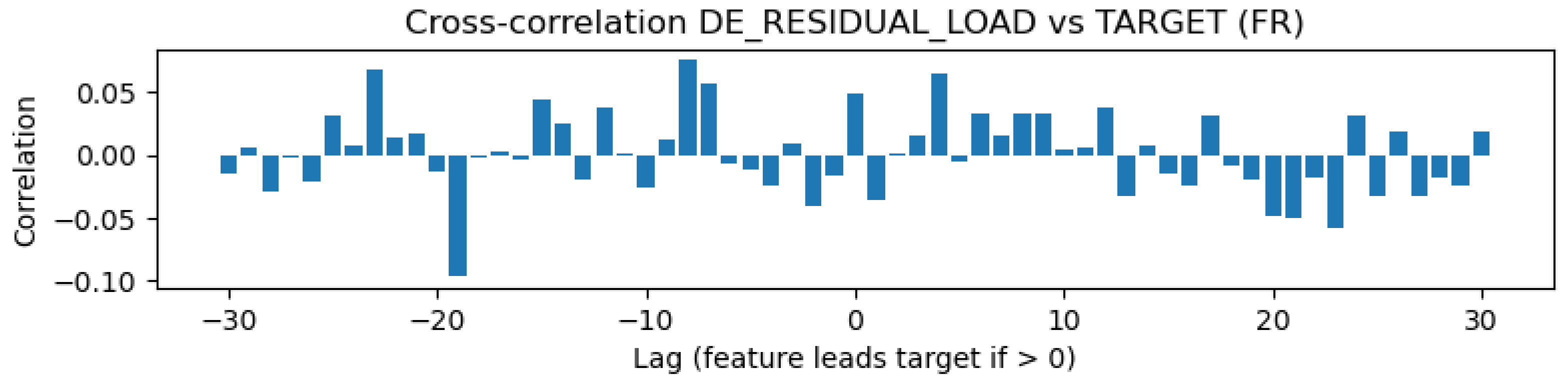

Autocorrelation of DE_HYDRO (FR)
ACF
0.1
0.0
0
5
10
15
20
25
30
Lag

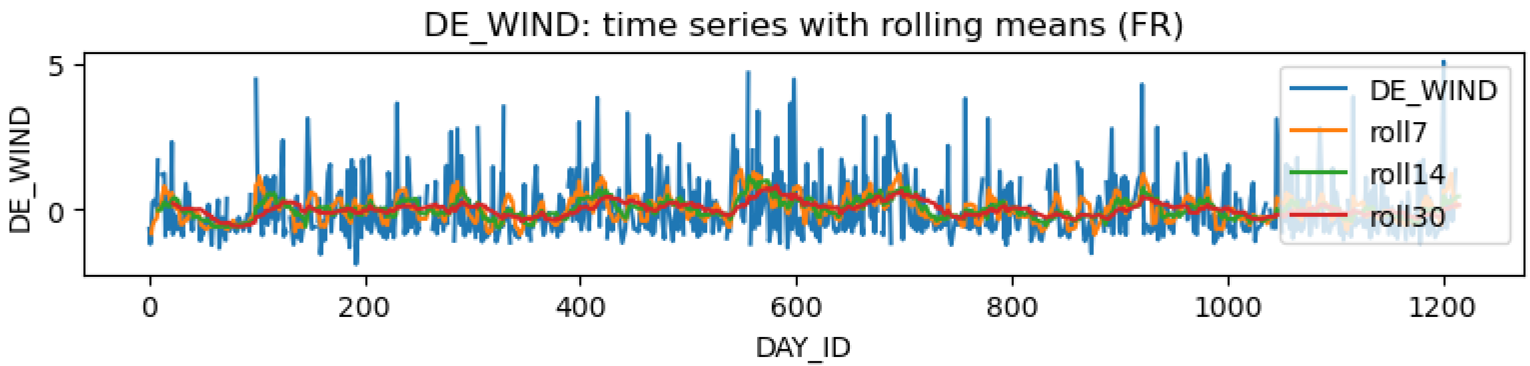

Cross-correlation DE_HYDRO vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−0.10
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

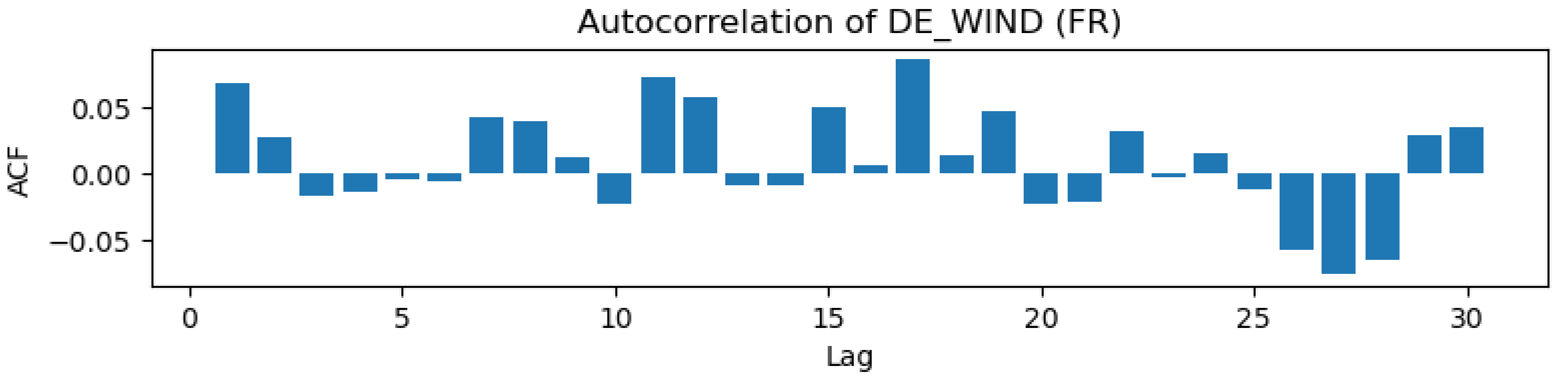

DE_CONSUMPTION: time series with rolling means (FR)
DE_CONSUMPTION
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_CONSUMPTION
roll7
roll14
roll30

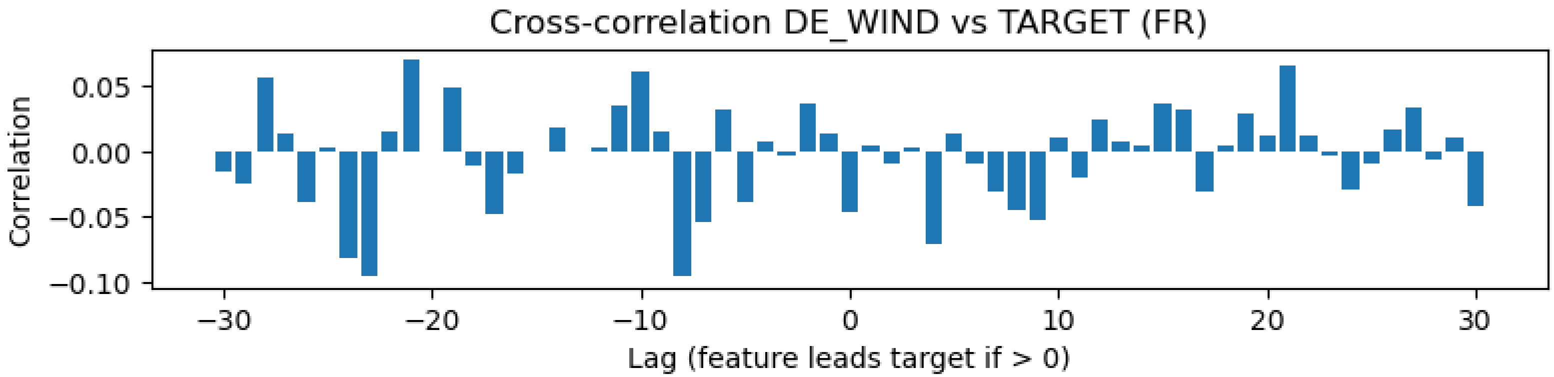

Autocorrelation of DE_CONSUMPTION (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag
Cross-correlation DE_CONSUMPTION vs TARGET (FR)
Correlation
0.10
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)
DE_RESIDUAL_LOAD: time series with rolling means (FR)
DE_RESIDUAL_LOAD
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_RESIDUAL_LOAD
roll7
roll14
roll30

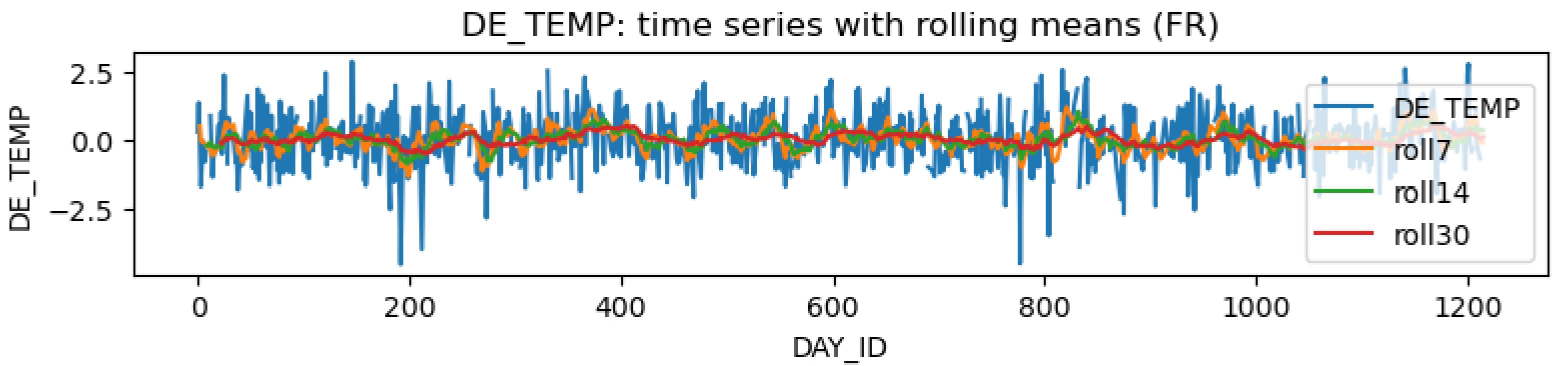

Autocorrelation of DE_RESIDUAL_LOAD (FR)
ACF
Lag

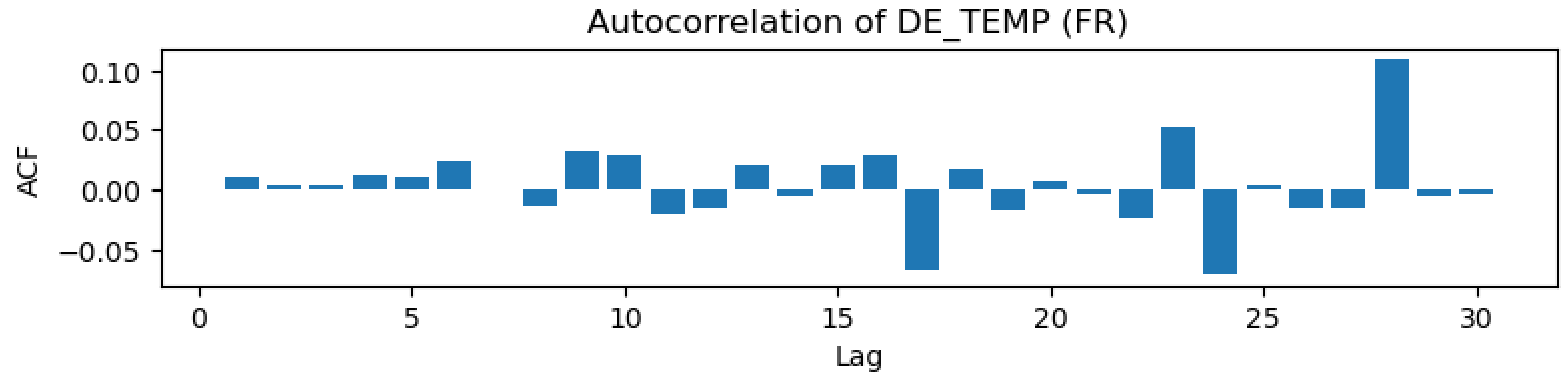

Cross-correlation DE_RESIDUAL_LOAD vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)


DE_WIND: time series with rolling means (FR)
DE_WIND
DAY_ID
DE_WIND
roll7
roll14
roll30

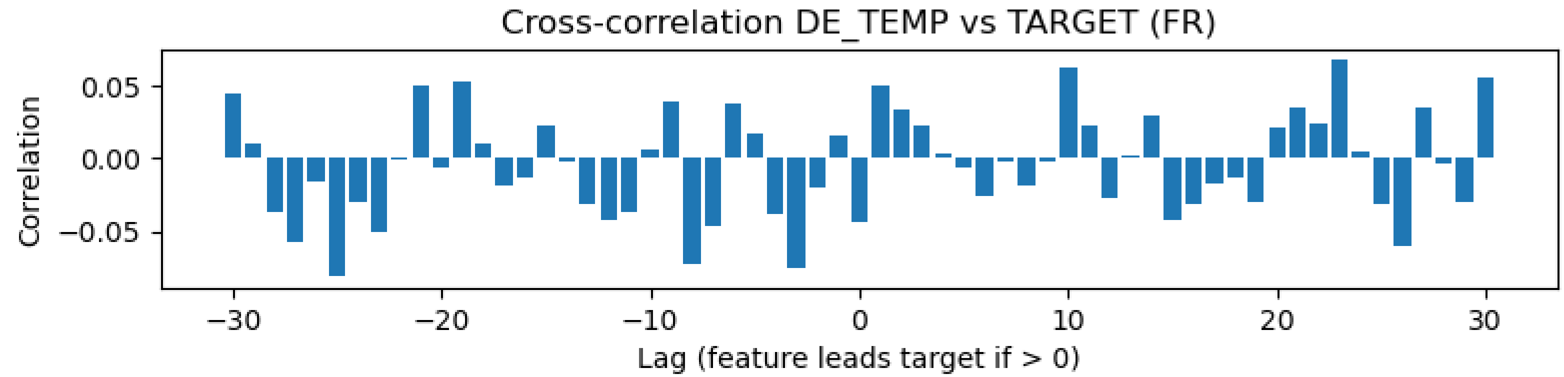

Autocorrelation of DE_WIND (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag


Cross-correlation DE_WIND vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−0.10
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

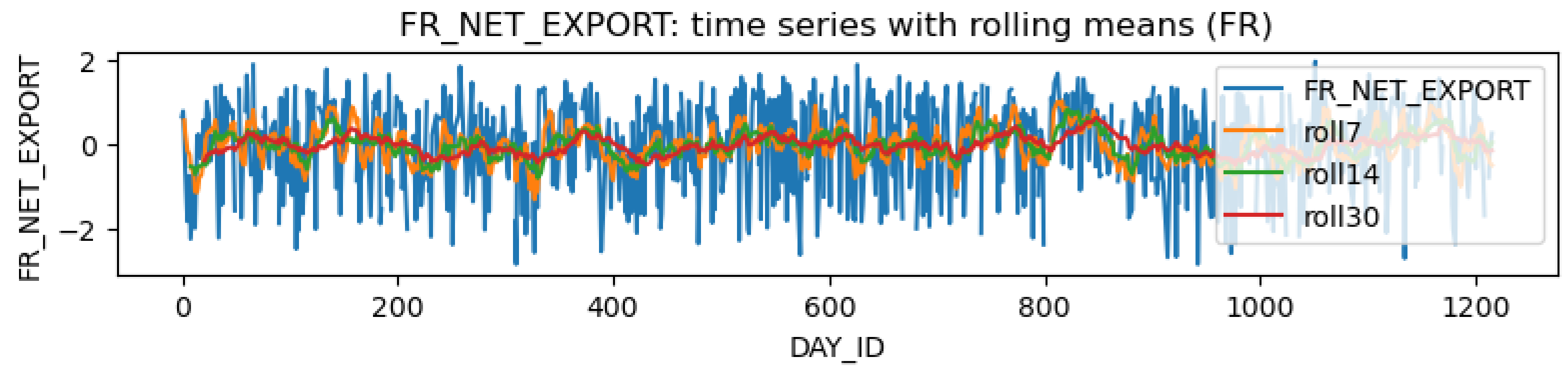

DE_TEMP: time series with rolling means (FR)
DE_TEMP
2.5
0.0
−2.5
0
200
400
600
800
1000
1200
DAY_ID
DE_TEMP
roll7
roll14
roll30



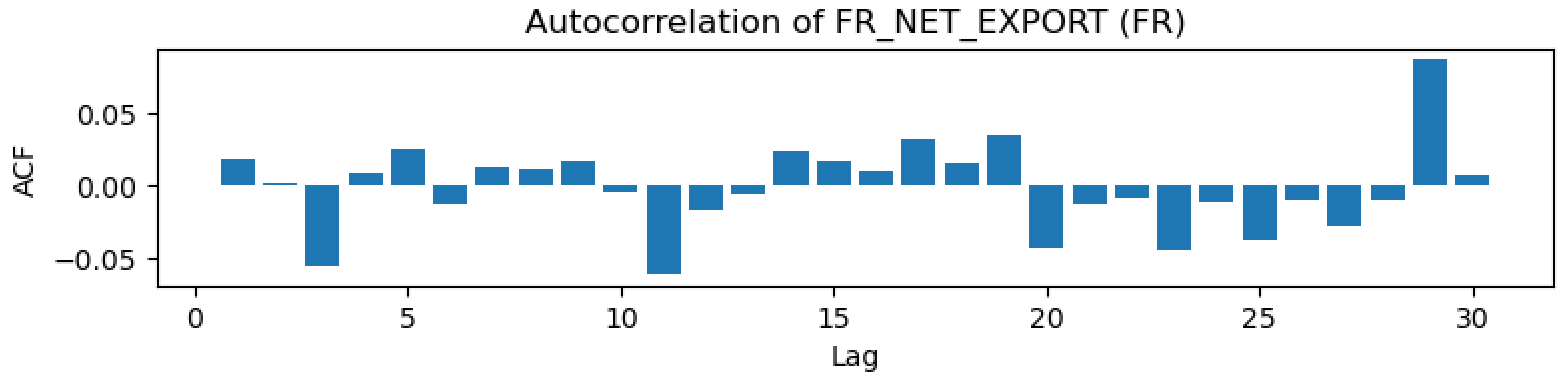
Autocorrelation of DE_TEMP (FR)
ACF
Lag

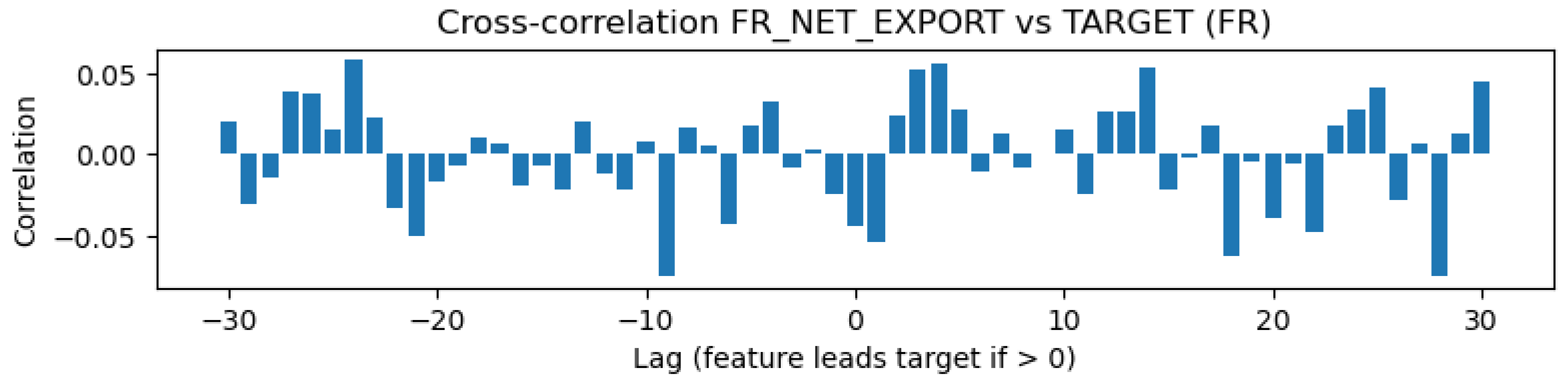


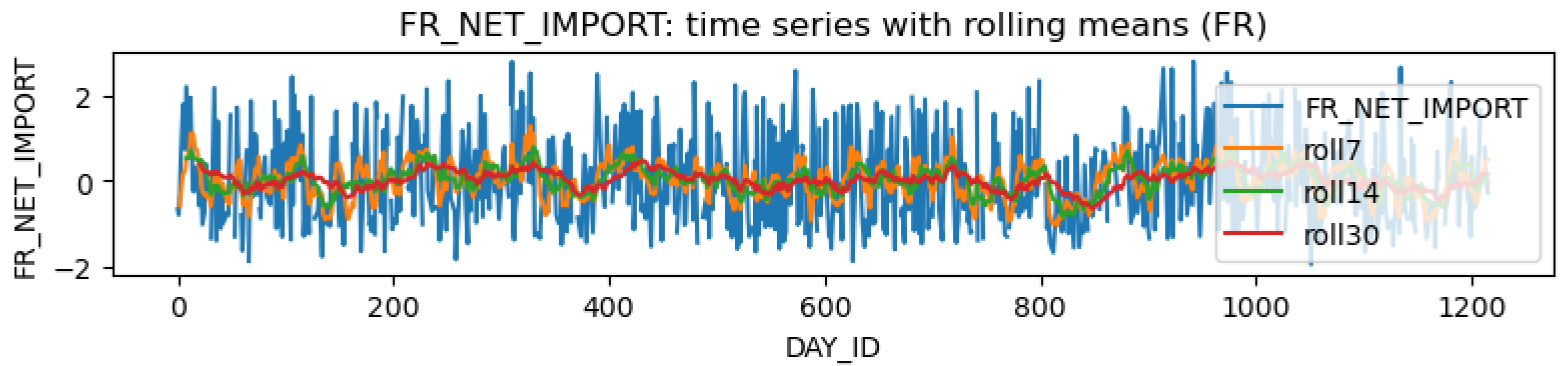
Cross-correlation DE_TEMP vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)

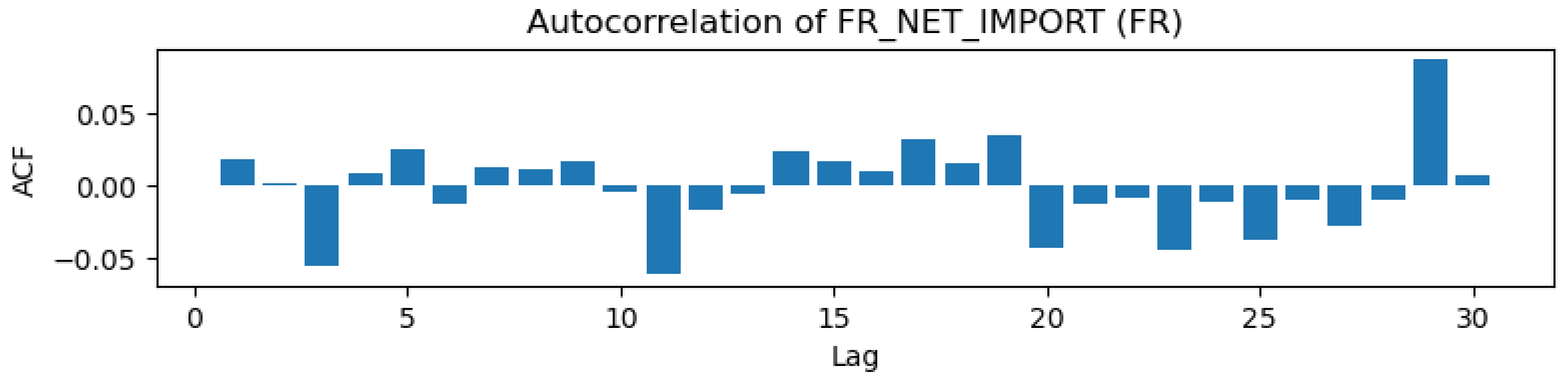


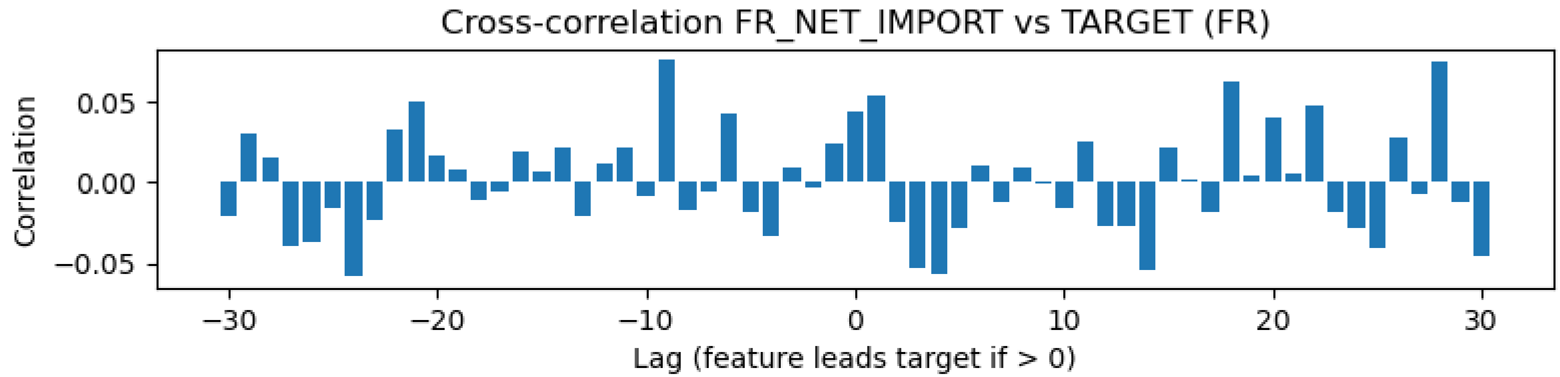
FR_NET_EXPORT: time series with rolling means (FR)
FR_NET_EXPORT
roll7
roll14
roll30
DAY_ID

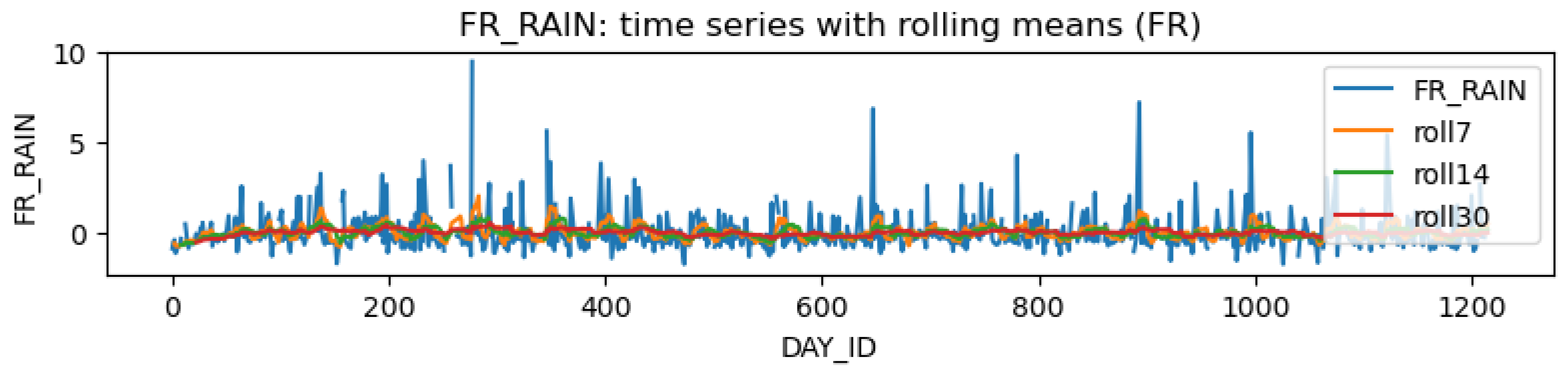

Autocorrelation of FR_NET_EXPORT (FR)
ACF
Lag

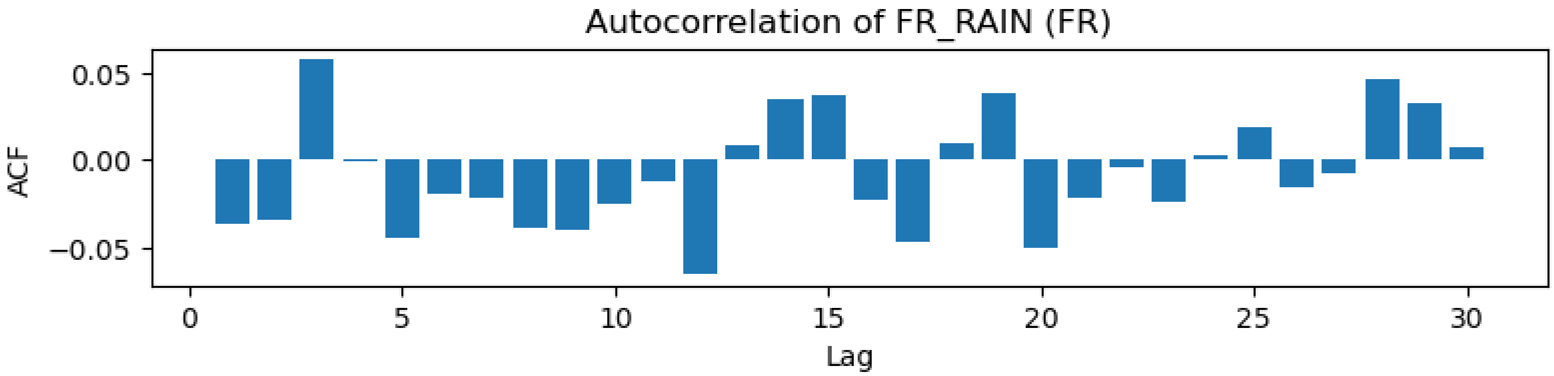

Cross-correlation FR_NET_EXPORT vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)

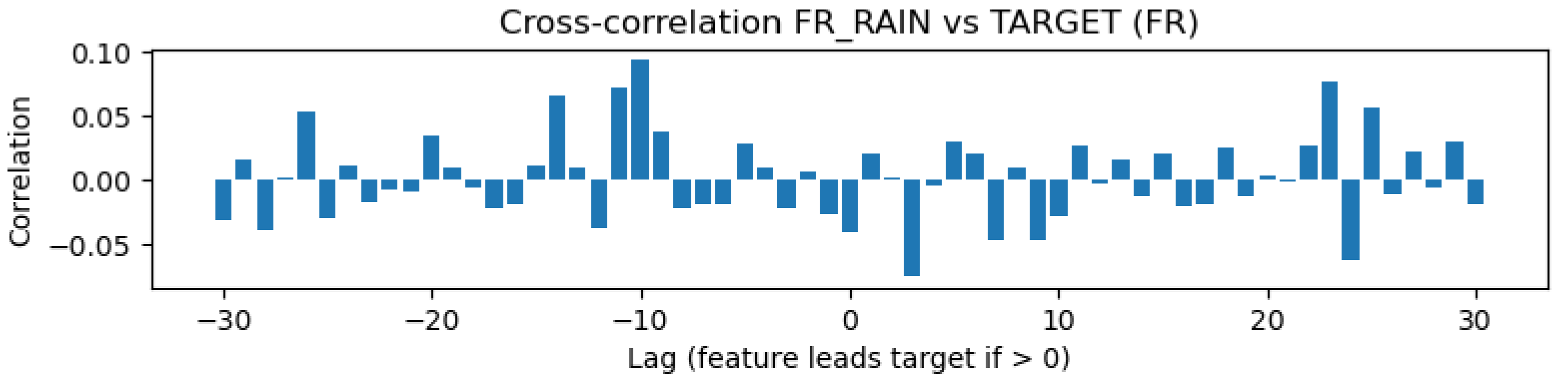

FR_NET_IMPORT: time series with rolling means (FR)
FR_NET_IMPORT
roll7
roll14
roll30
DAY_ID

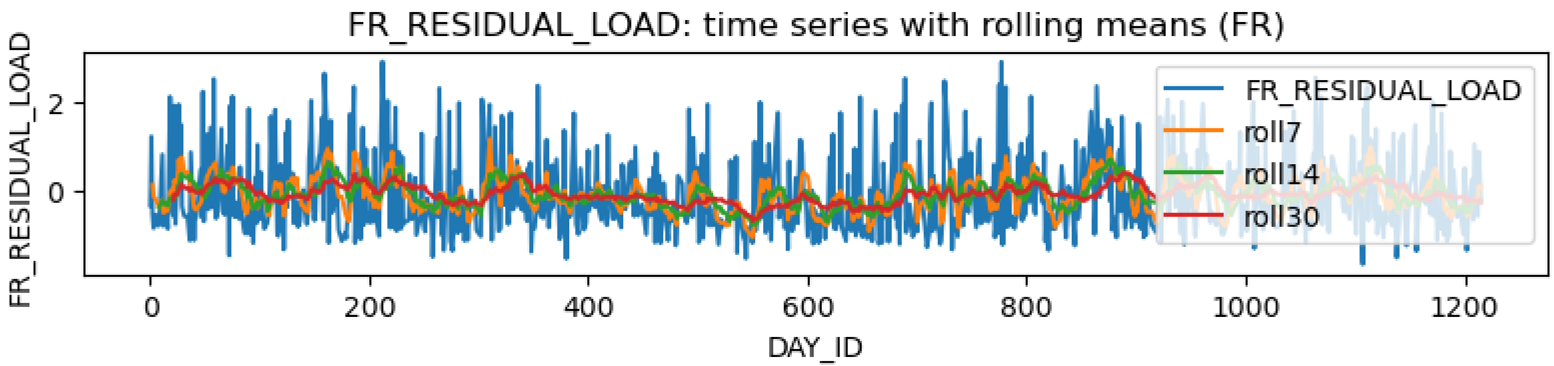

Autocorrelation of FR_NET_IMPORT (FR)
ACF
Lag

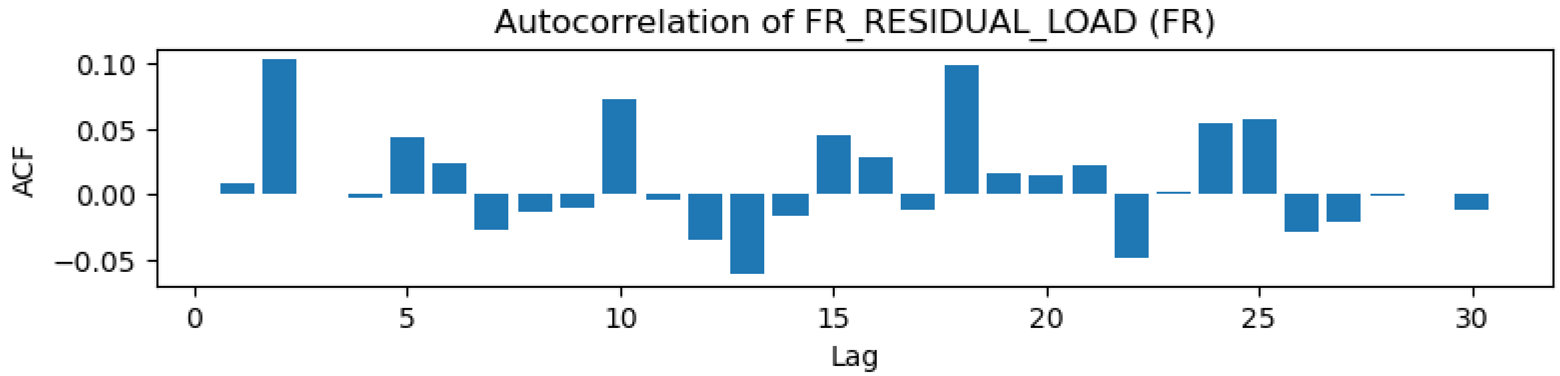

Cross-correlation FR_NET_IMPORT vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)

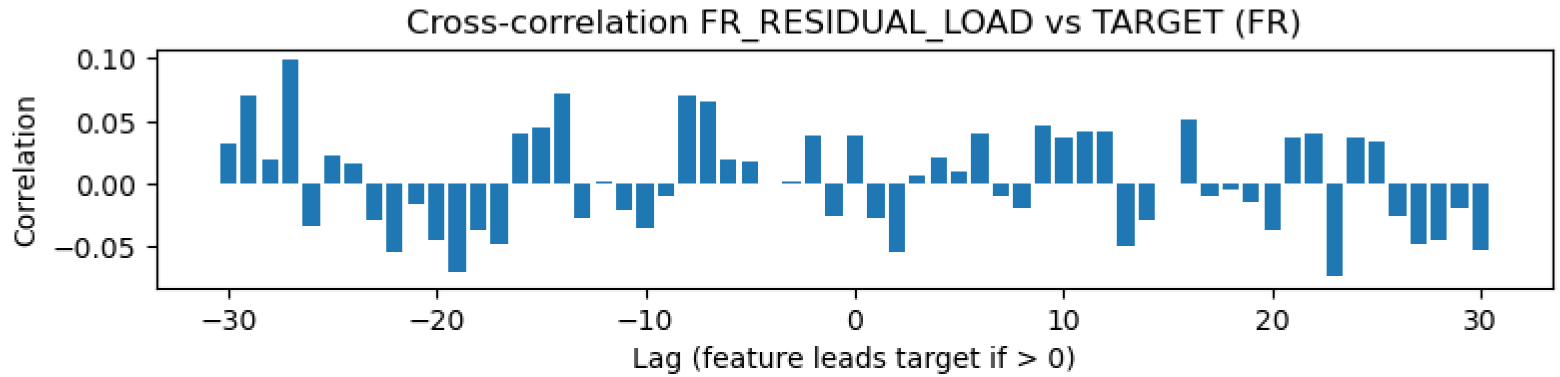

FR_RAIN: time series with rolling means (FR)
FR_RAIN
roll7
roll14
roll30
DAY_ID

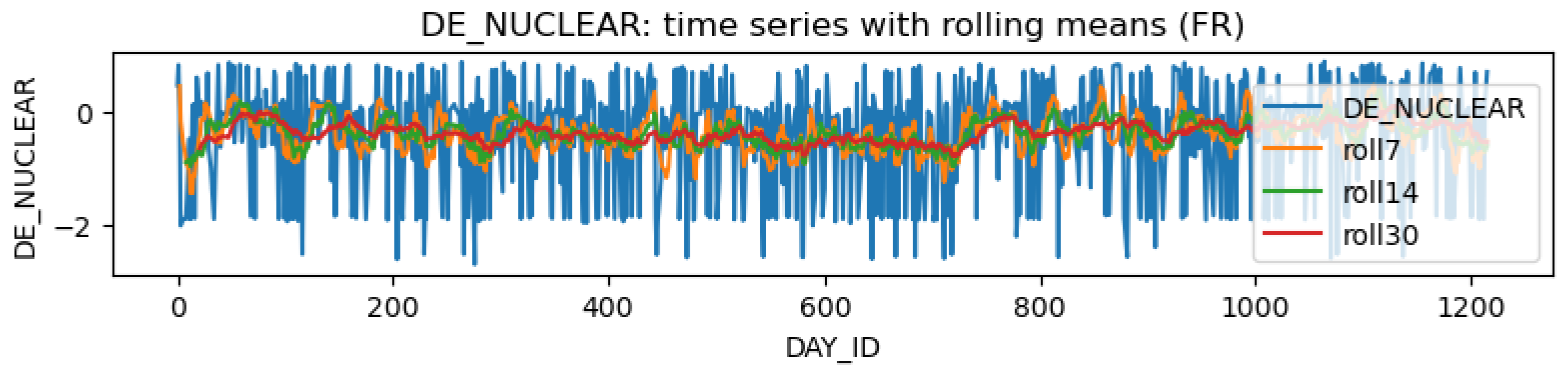

Autocorrelation of FR_RAIN (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

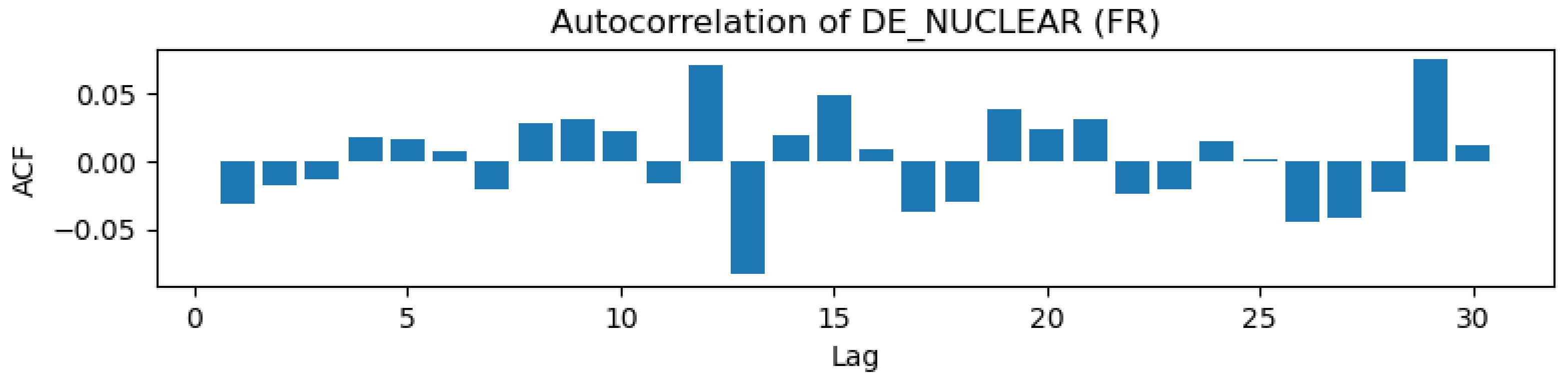

Cross-correlation FR_RAIN vs TARGET (FR)
Correlation
0.10
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

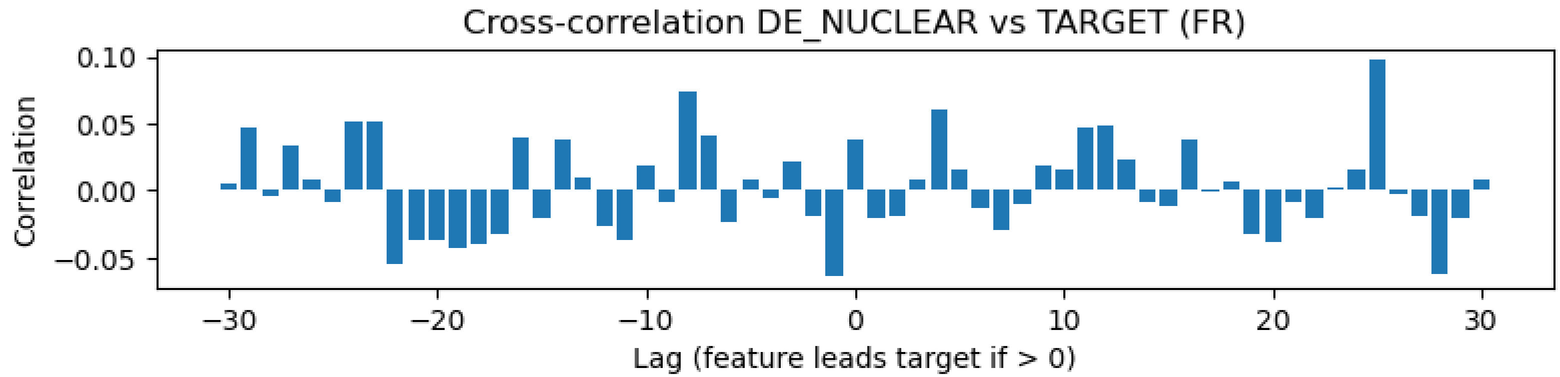

FR_RESIDUAL_LOAD: time series with rolling means (FR)
FR_RESIDUAL_LOAD
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_RESIDUAL_LOAD
roll7
roll14
roll30

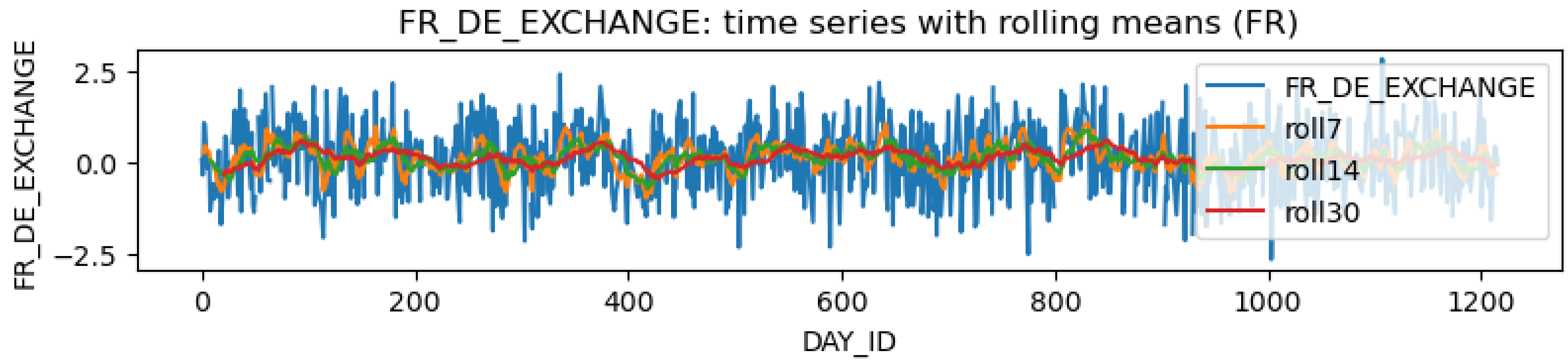

Autocorrelation of FR_RESIDUAL_LOAD (FR)
ACF
Lag

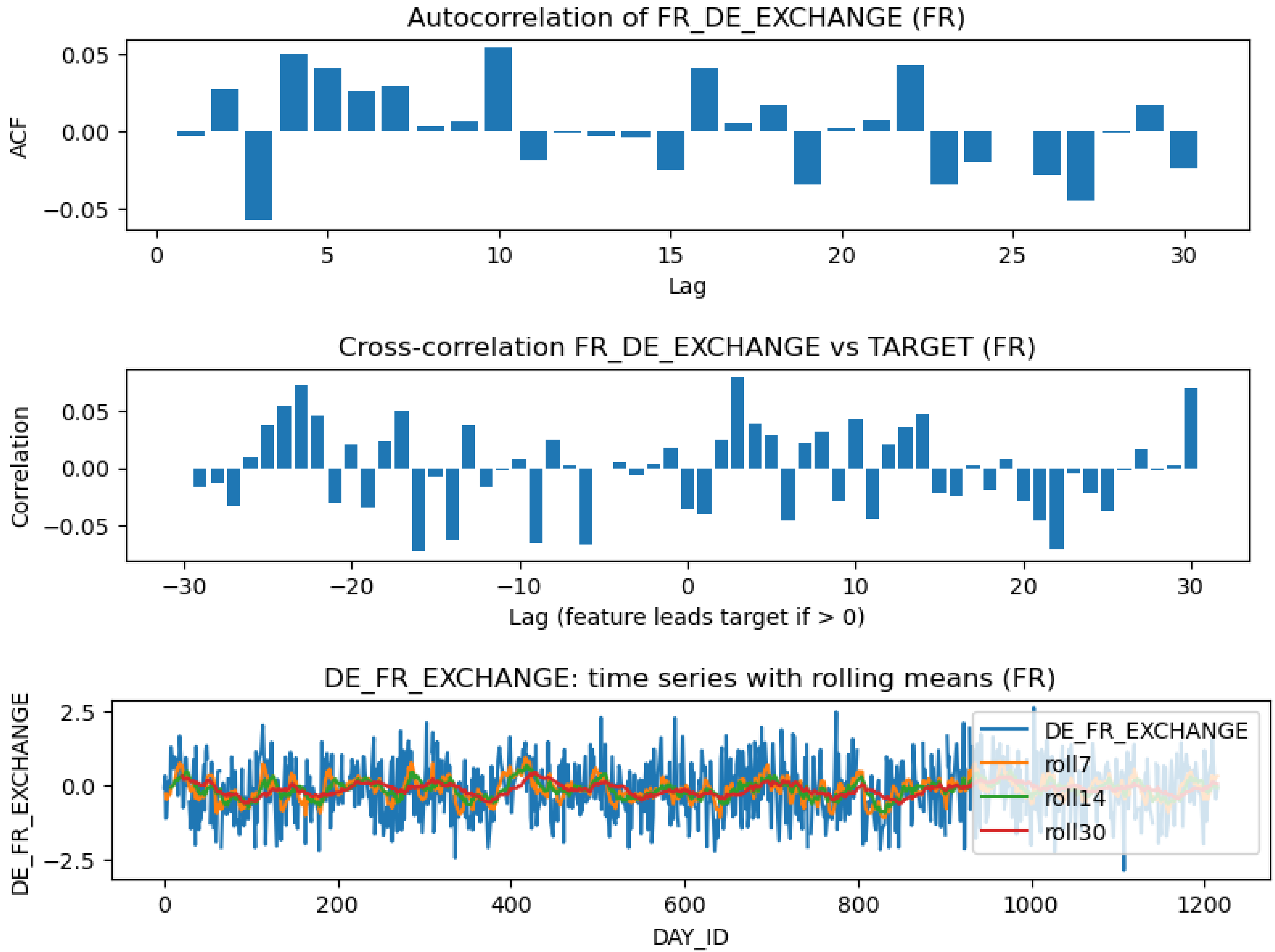

Cross-correlation FR_RESIDUAL_LOAD vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)


DE_NUCLEAR: time series with rolling means (FR)
DE_NUCLEAR
roll7
roll14
roll30
DAY_ID

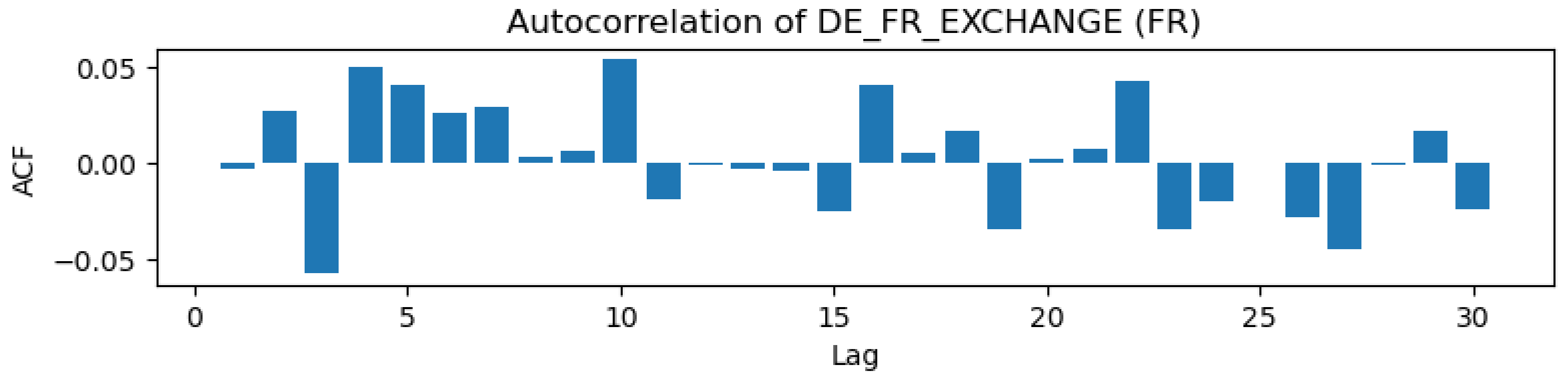

Autocorrelation of DE_NUCLEAR (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

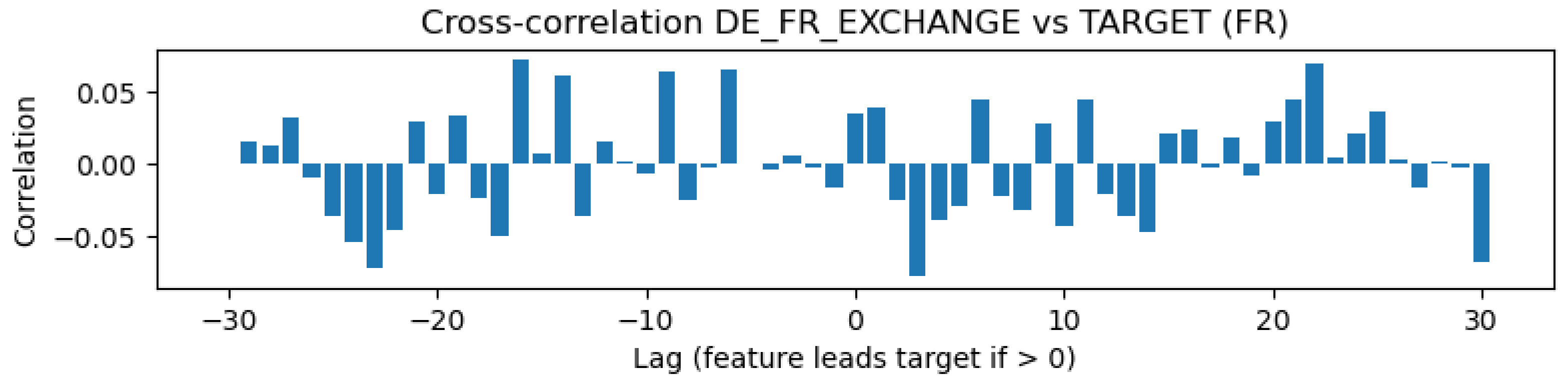

Cross-correlation DE_NUCLEAR vs TARGET (FR)
Correlation
0.10
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

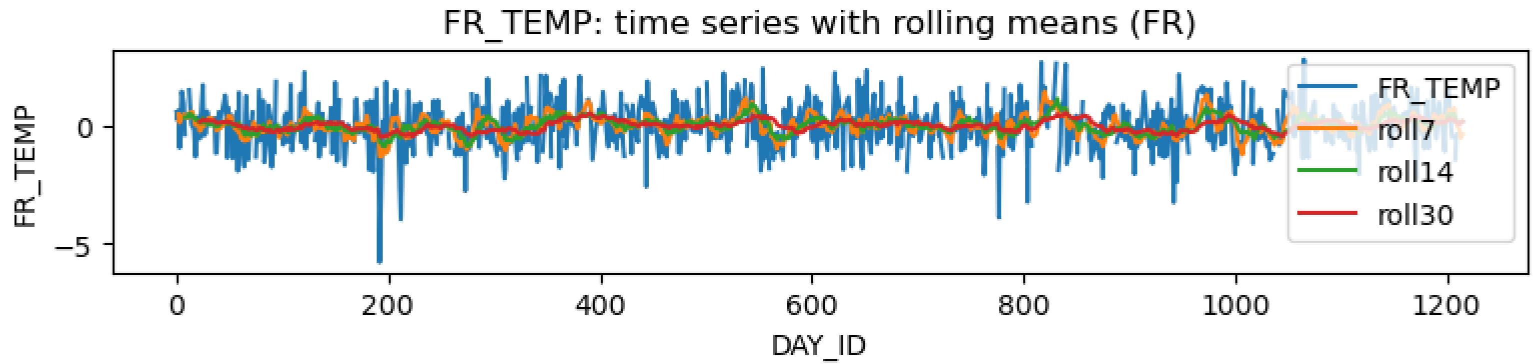

FR_DE_EXCHANGE: time series with rolling means (FR)
FR_DE_EXCHANGE
2.5
0.0
−2.5
0
200
400
600
800
1000
1200
DAY_ID
FR_DE_EXCHANGE
roll7
roll14
roll30

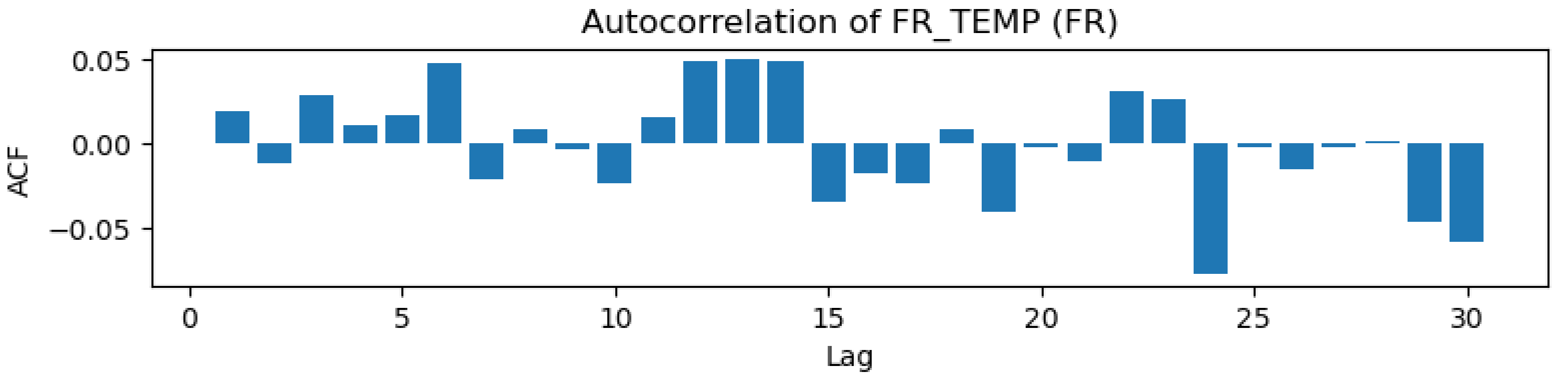


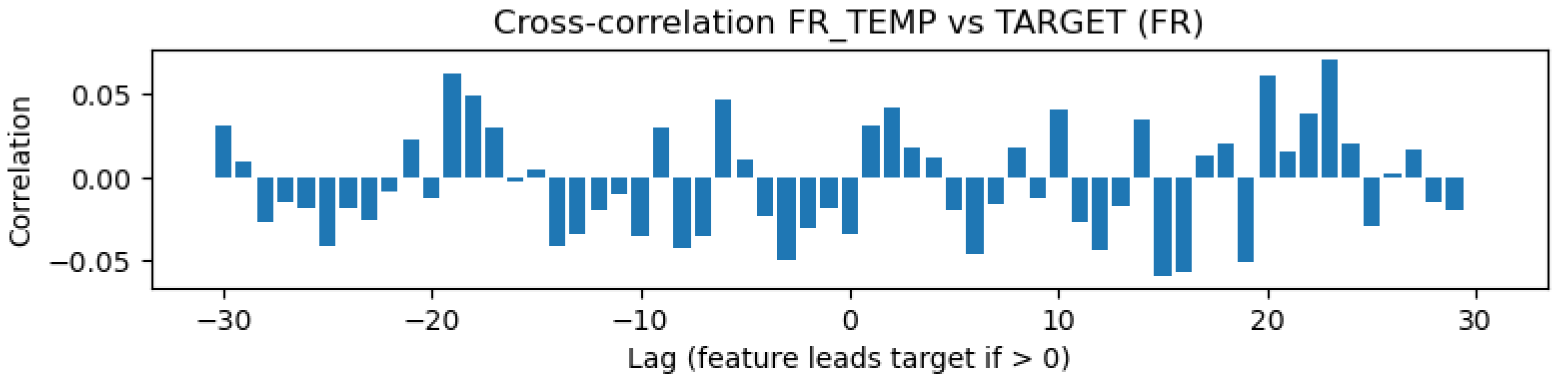
Autocorrelation of FR_DE_EXCHANGE (FR)
ACF
Lag
Cross-correlation FR_DE_EXCHANGE vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)
DE_FR_EXCHANGE: time series with rolling means (FR)
DE_FR_EXCHANGE
roll7
roll14
roll30
DAY_ID

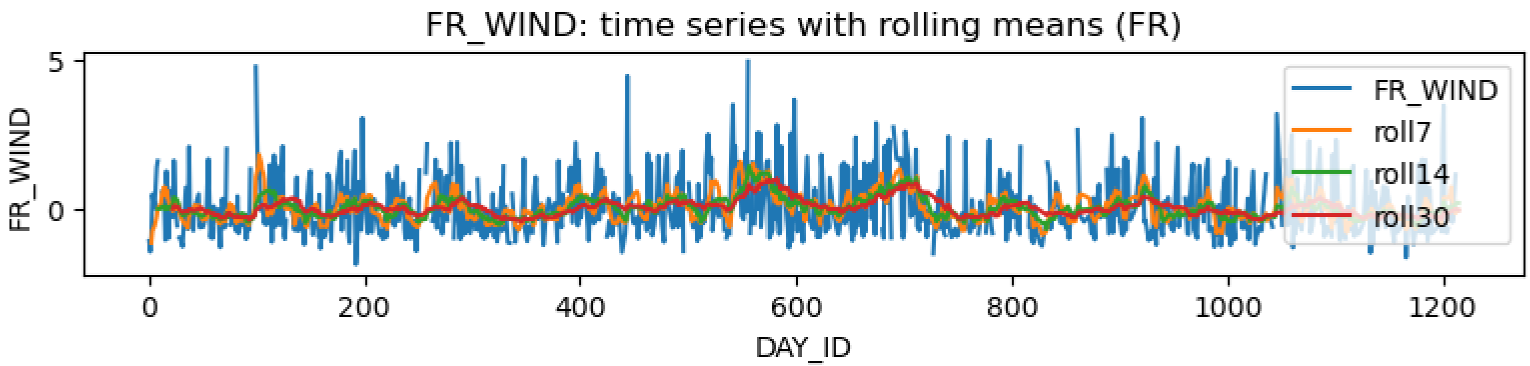

Autocorrelation of DE_FR_EXCHANGE (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

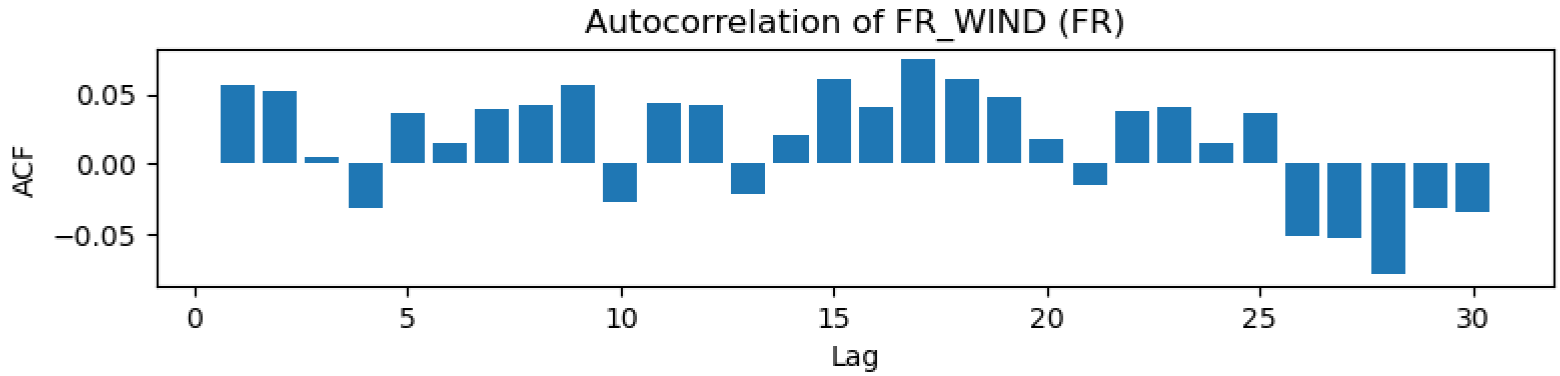

Cross-correlation DE_FR_EXCHANGE vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

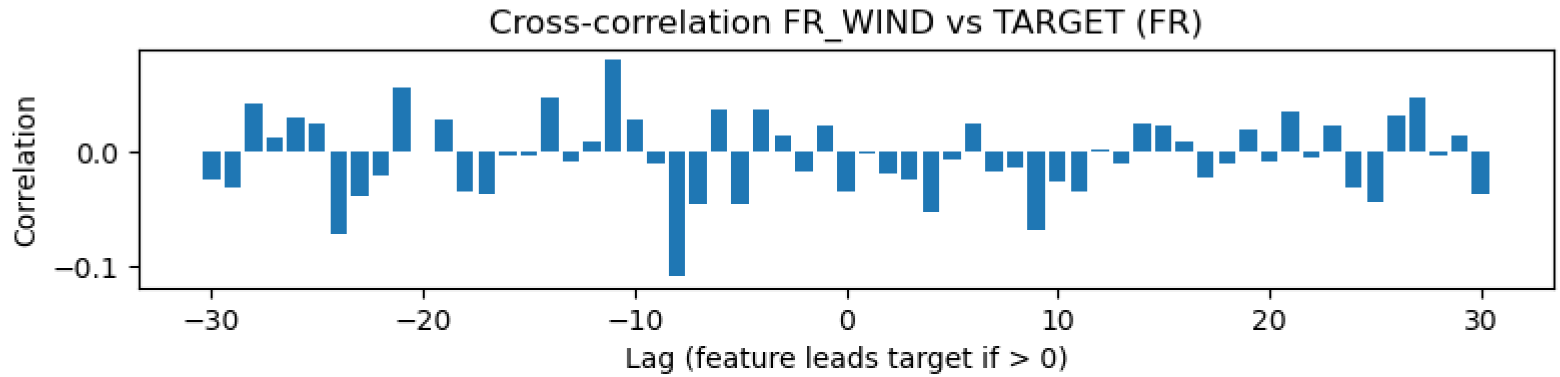

FR_TEMP: time series with rolling means (FR)
FR_TEMP
0
−5
0
200
400
600
800
1000
1200
DAY_ID
FR_TEMP
roll7
roll14
roll30

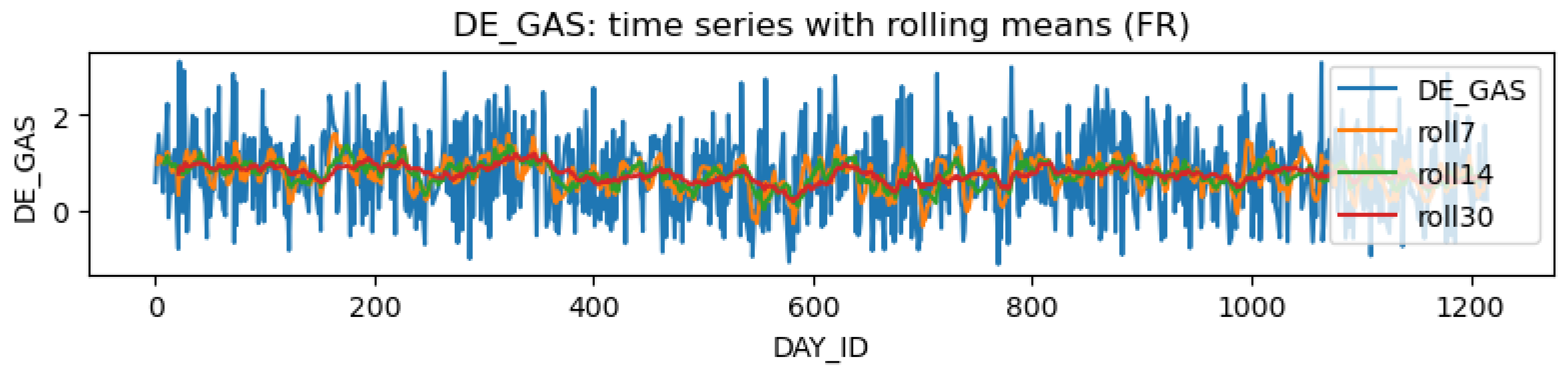

Autocorrelation of FR_TEMP (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

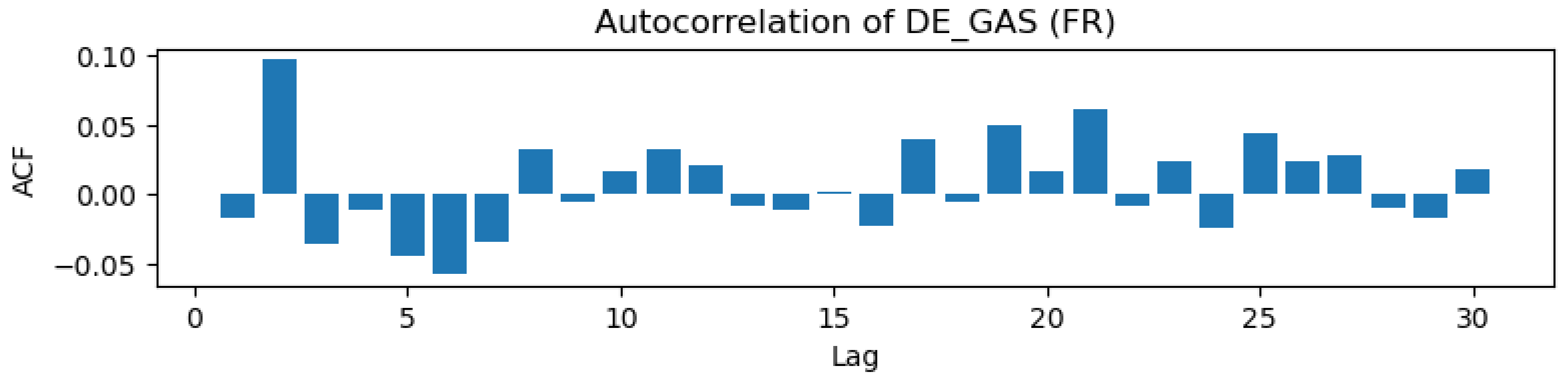

Cross-correlation FR_TEMP vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

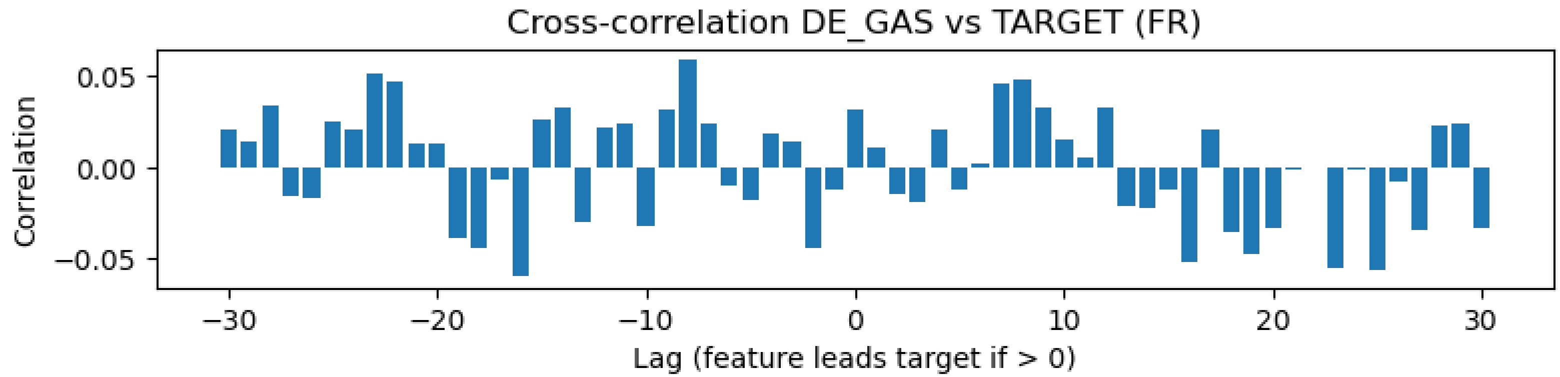

FR_WIND: time series with rolling means (FR)
FR_WIND
5
0
0
200
400
600
800
1000
1200
DAY_ID
FR_WIND
roll7
roll14
roll30

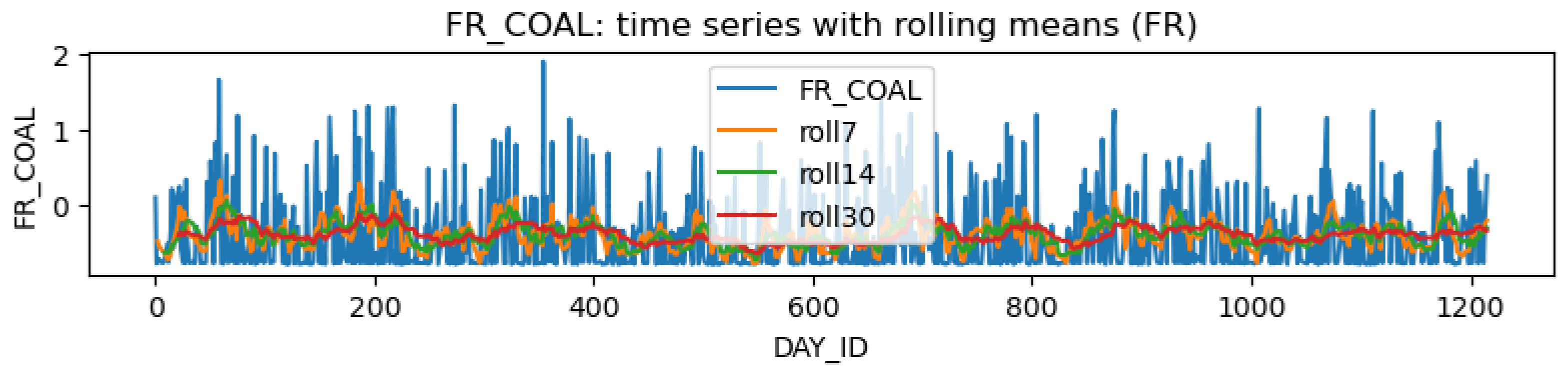

Autocorrelation of FR_WIND (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

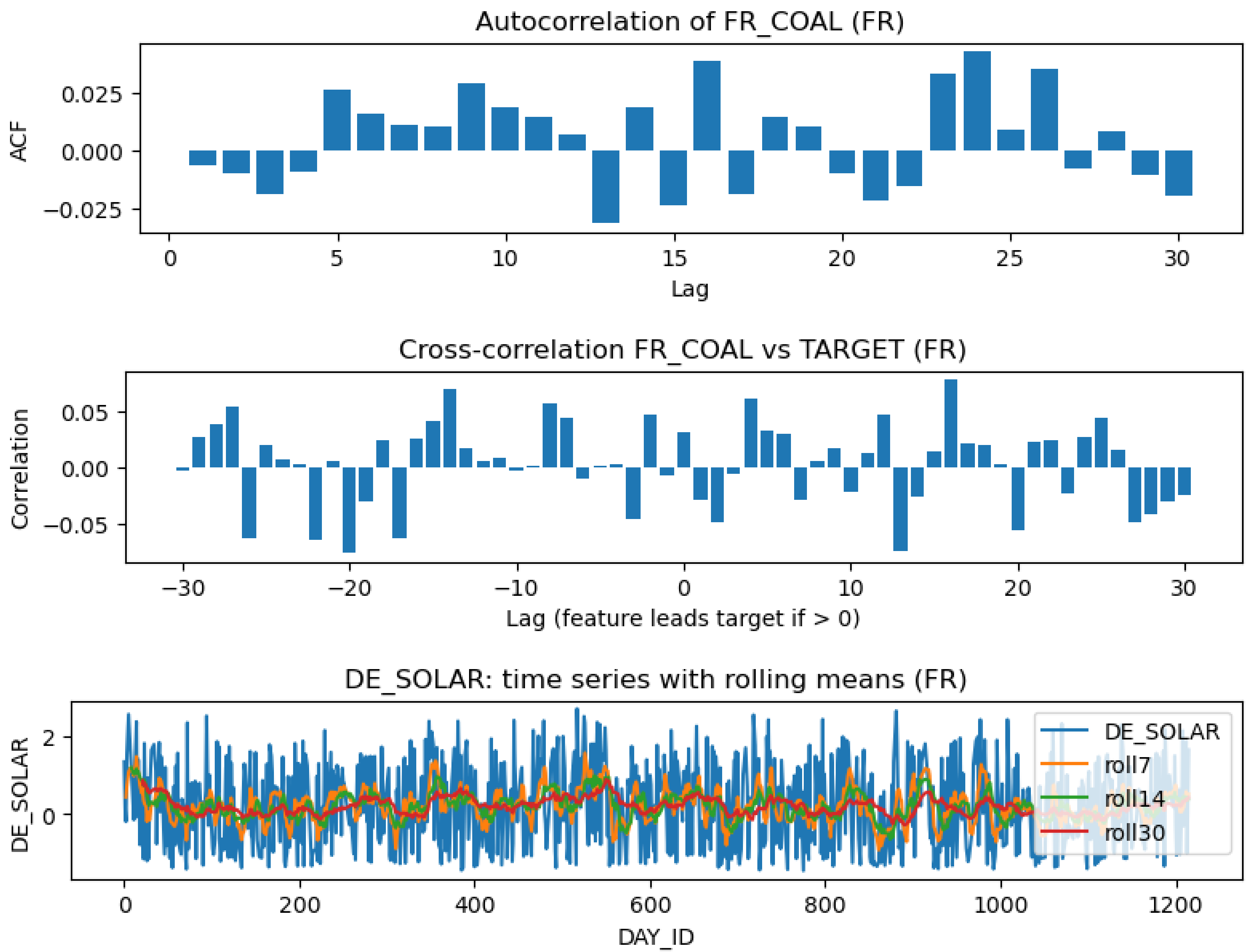

Cross-correlation FR_WIND vs TARGET (FR)
Correlation
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

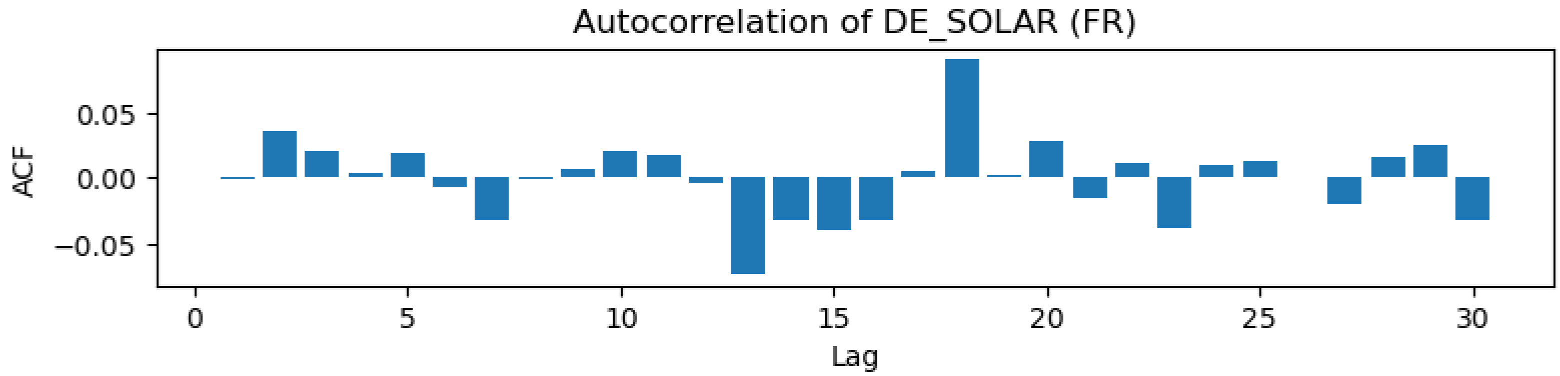

DE_GAS: time series with rolling means (FR)
DE_GAS
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_GAS
roll7
roll14
roll30

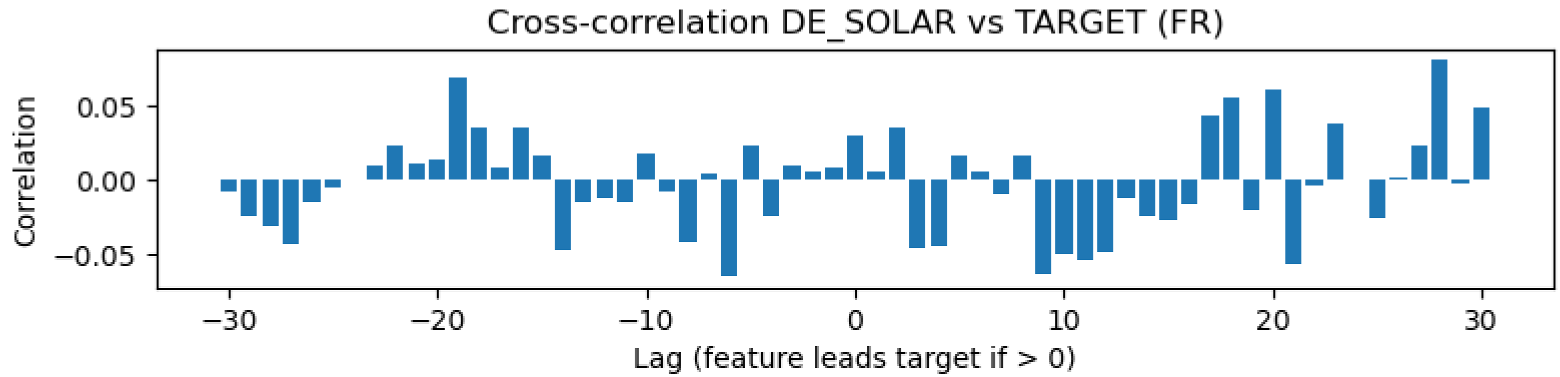

Autocorrelation of DE_GAS (FR)
ACF
0.10
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

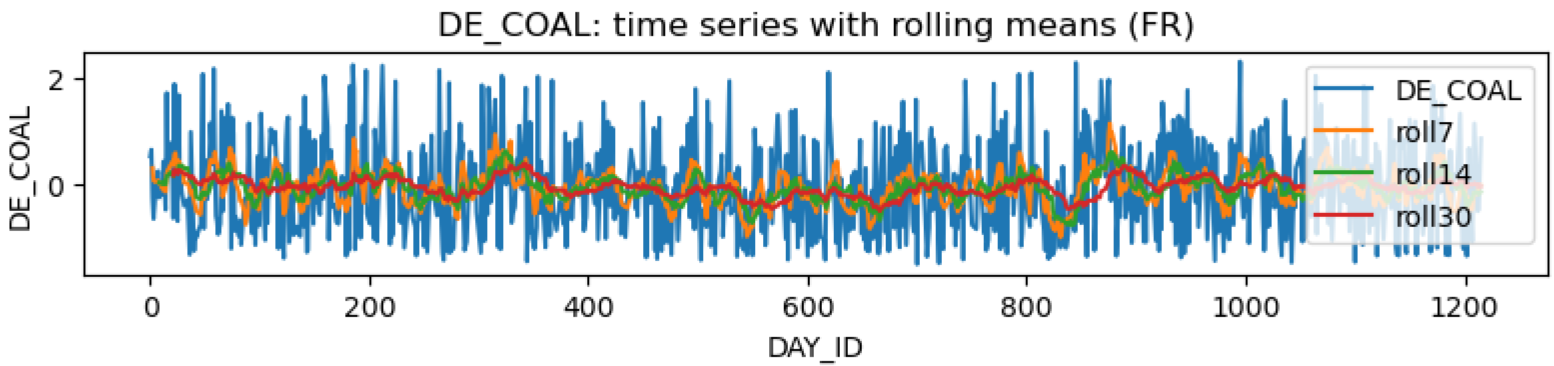

Cross-correlation DE_GAS vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

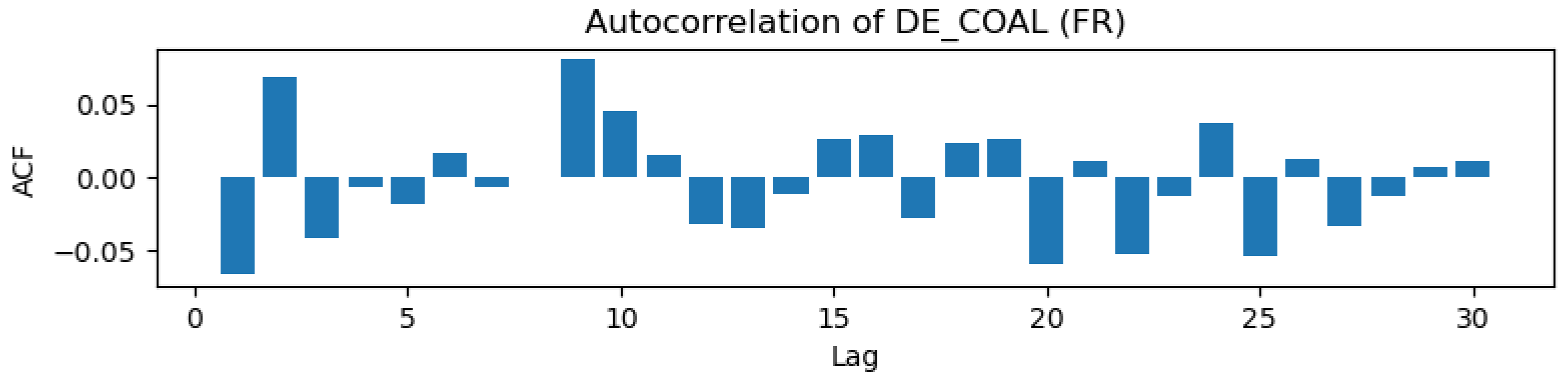

FR_COAL: time series with rolling means (FR)
FR_COAL
2
1
0
FR_COAL
roll7
roll14
roll30
0
200
400
600
800
1000
1200
DAY_ID

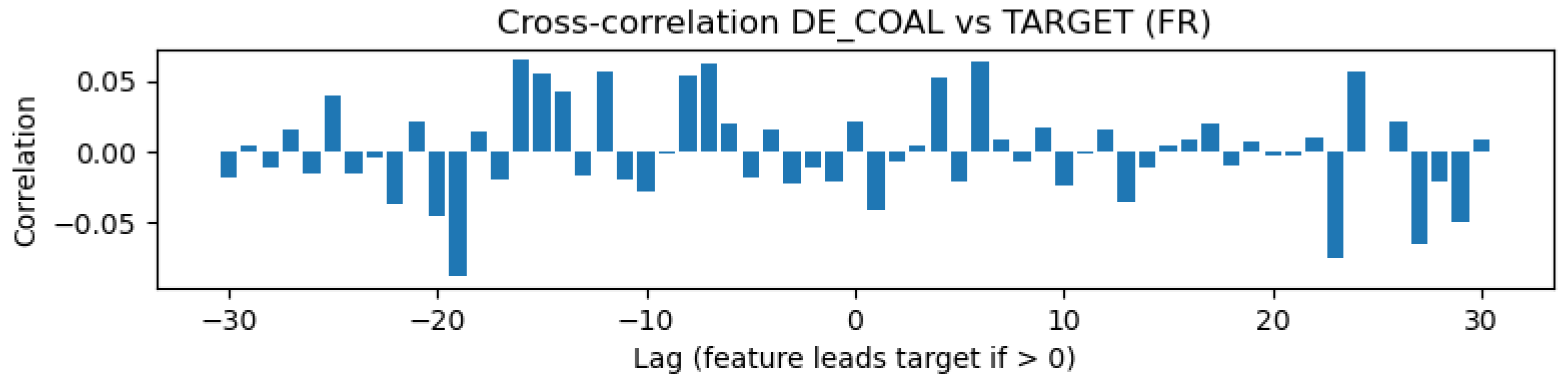

Autocorrelation of FR_COAL (FR)
ACF
Lag
Cross-correlation FR_COAL vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)
DE_SOLAR: time series with rolling means (FR)
DE_SOLAR
roll7
roll14
roll30
DAY_ID

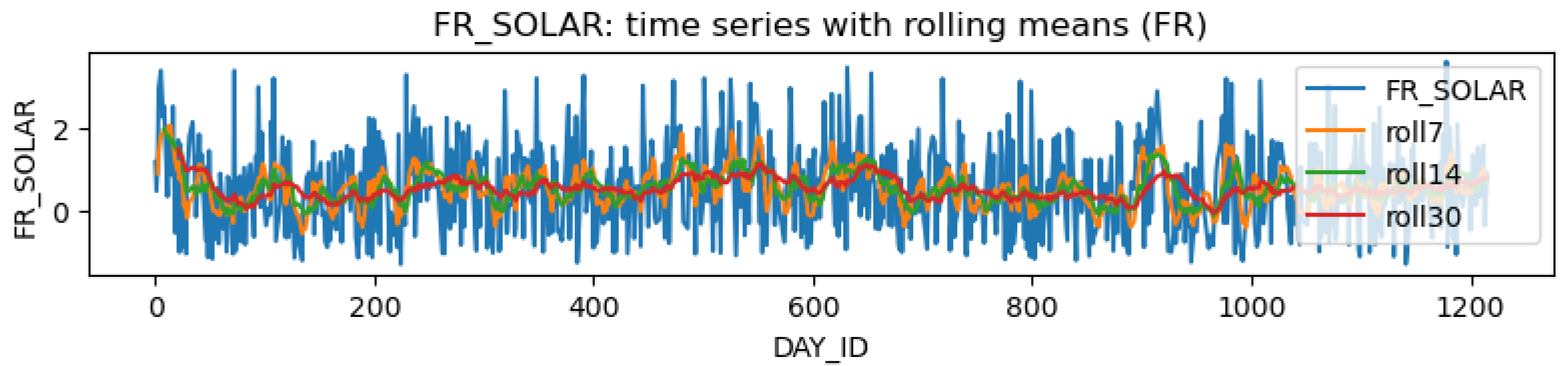

Autocorrelation of DE_SOLAR (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

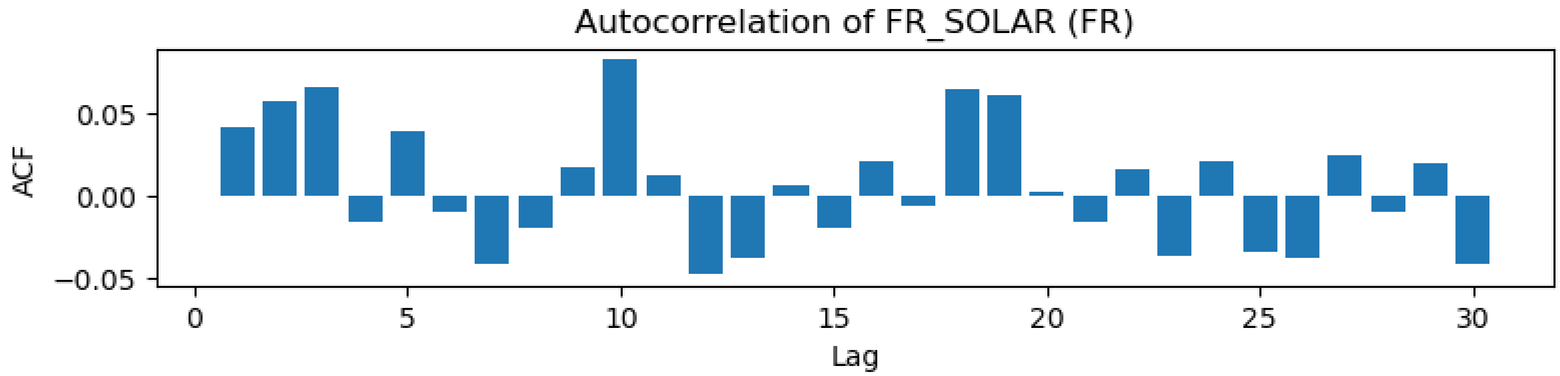

Cross-correlation DE_SOLAR vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

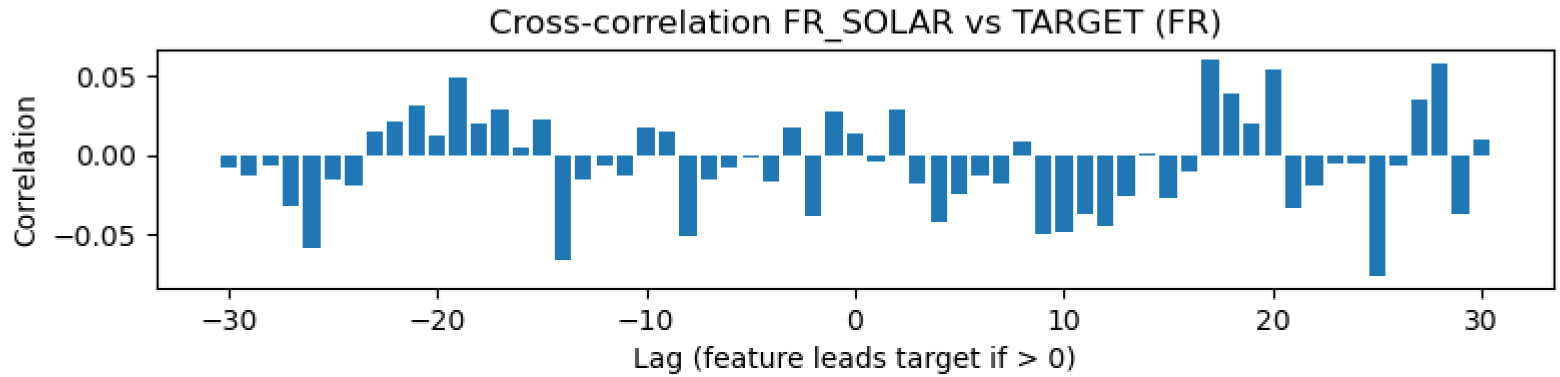

DE_COAL: time series with rolling means (FR)
DE_COAL
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_COAL
roll7
roll14
roll30

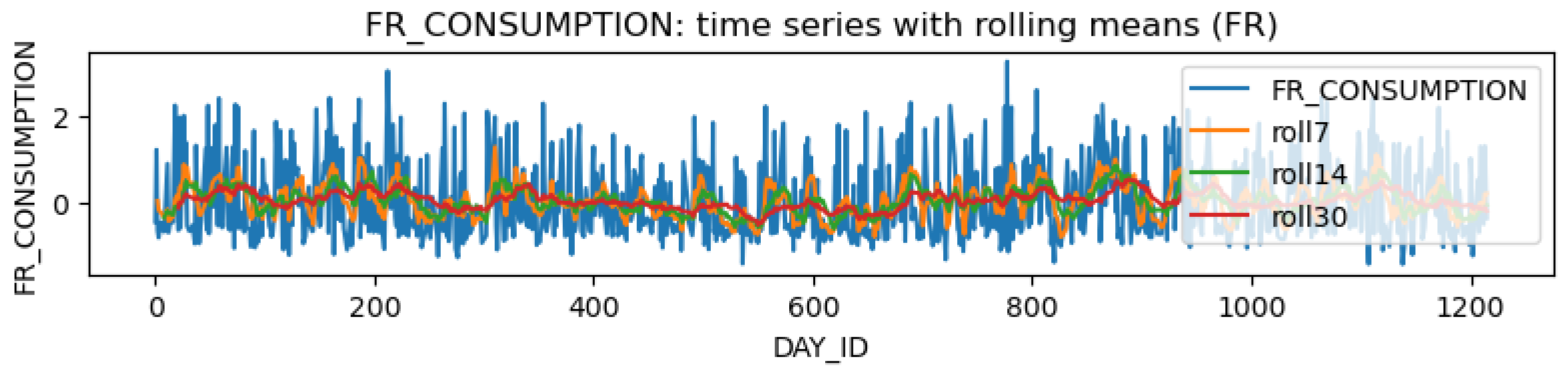

Autocorrelation of DE_COAL (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

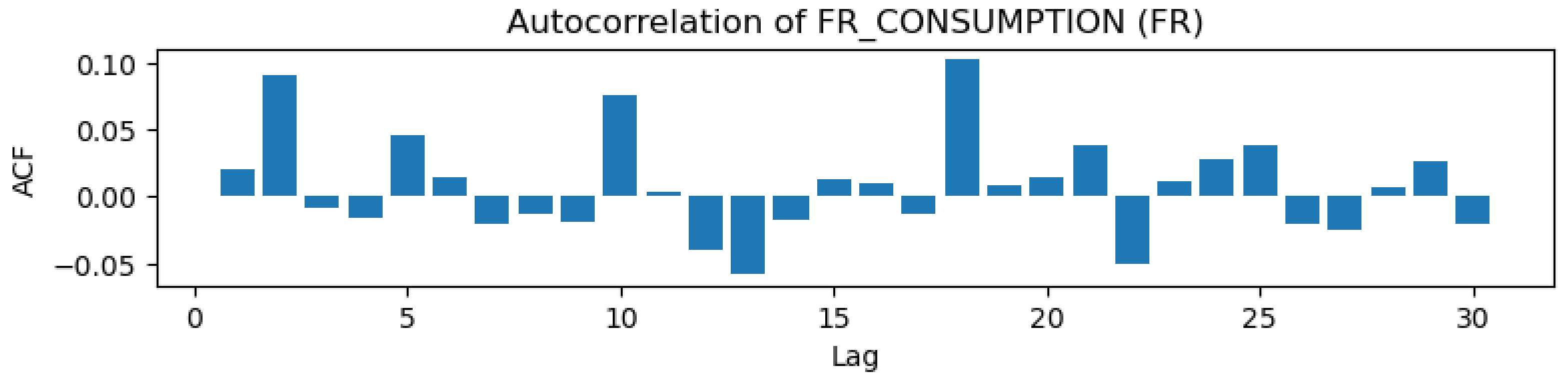

Cross-correlation DE_COAL vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

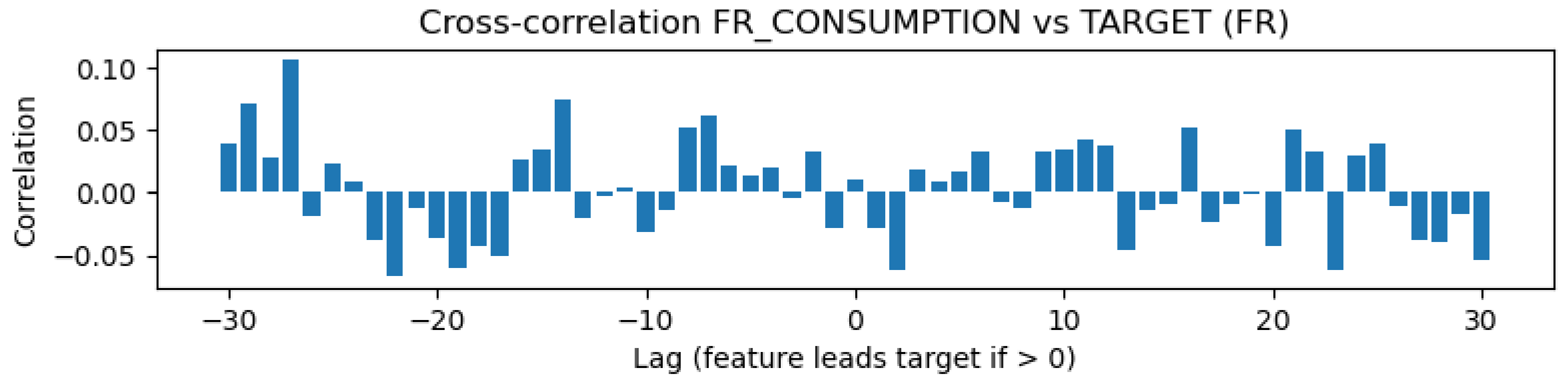

FR_SOLAR: time series with rolling means (FR)
FR_SOLAR
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_SOLAR
roll7
roll14
roll30

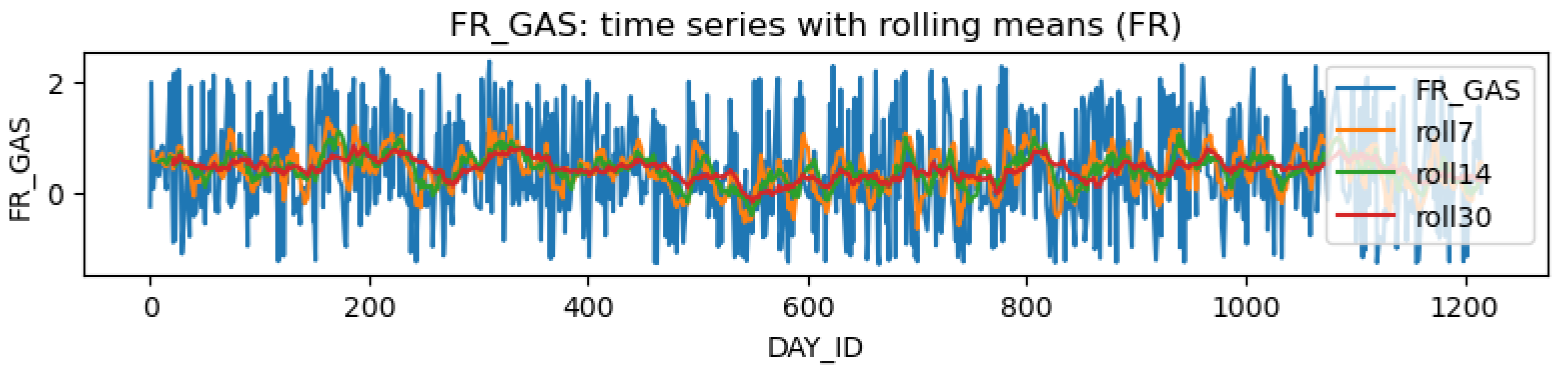

Autocorrelation of FR_SOLAR (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

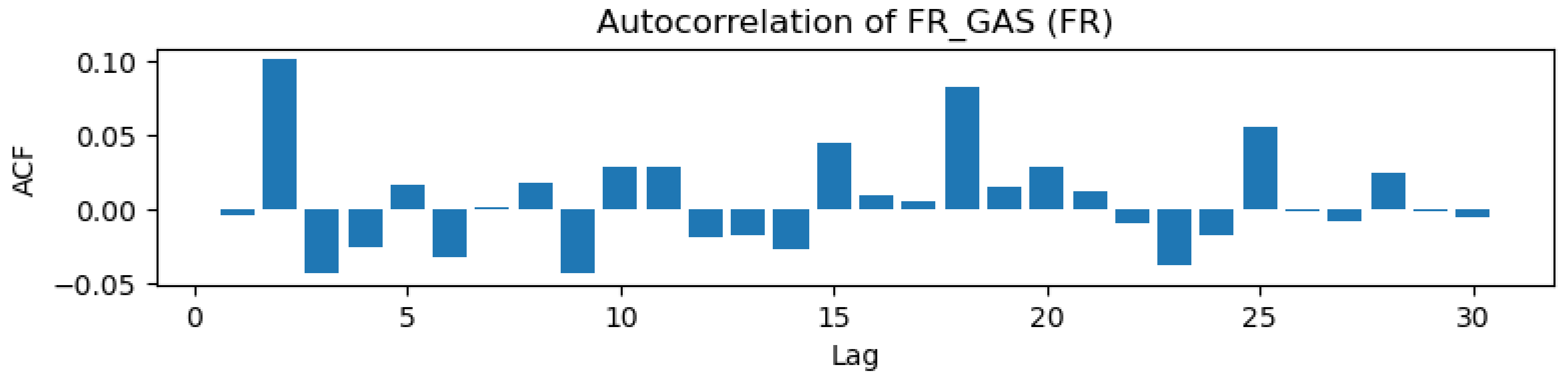

Cross-correlation FR_SOLAR vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

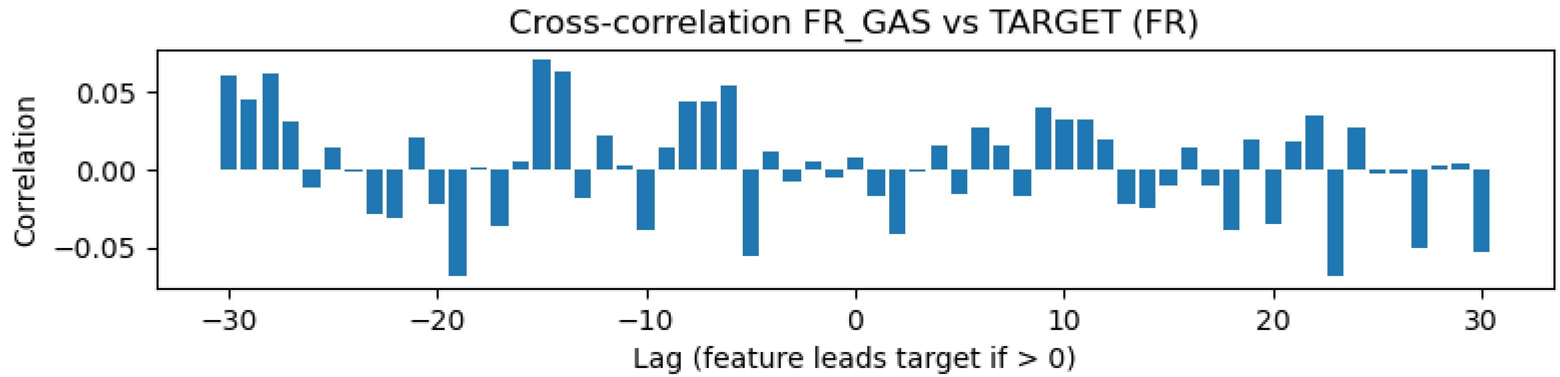

FR_CONSUMPTION: time series with rolling means (FR)
FR_CONSUMPTION
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_CONSUMPTION
roll7
roll14
roll30

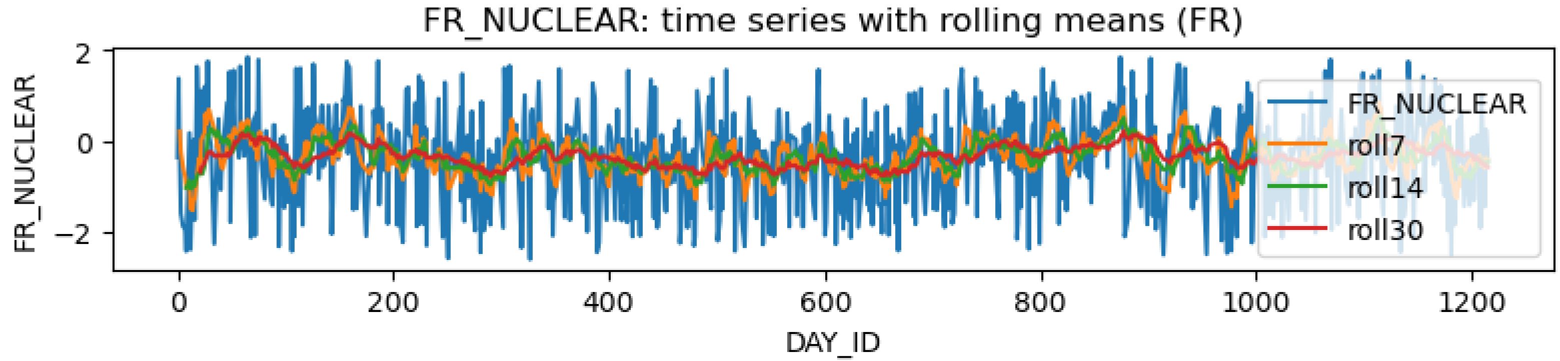

Autocorrelation of FR_CONSUMPTION (FR)
ACF
Lag
0.10
0.05
0.00
−0.05
0
5
10
15
20
25
30

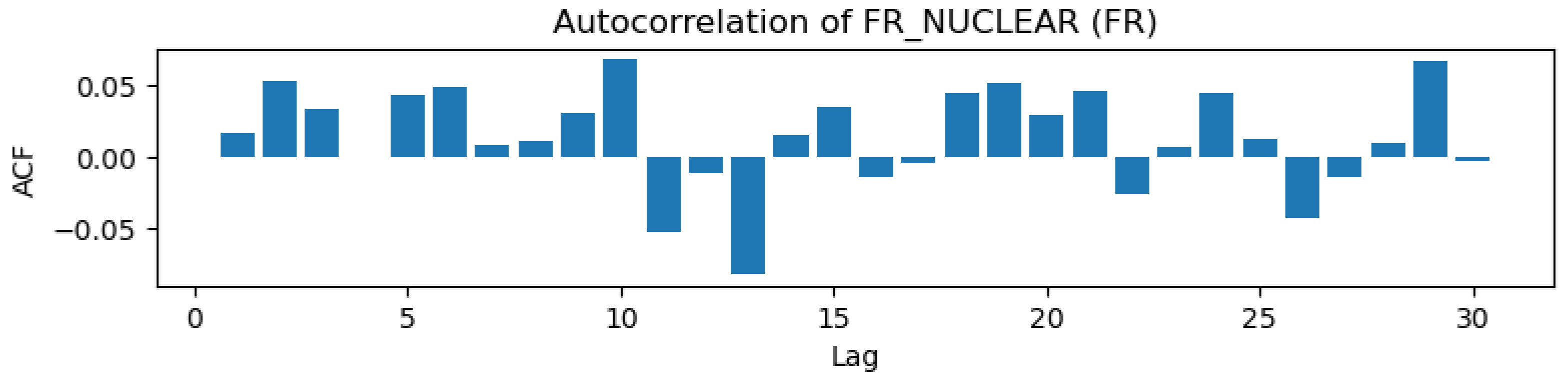

Cross-correlation FR_CONSUMPTION vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)
0.10
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30

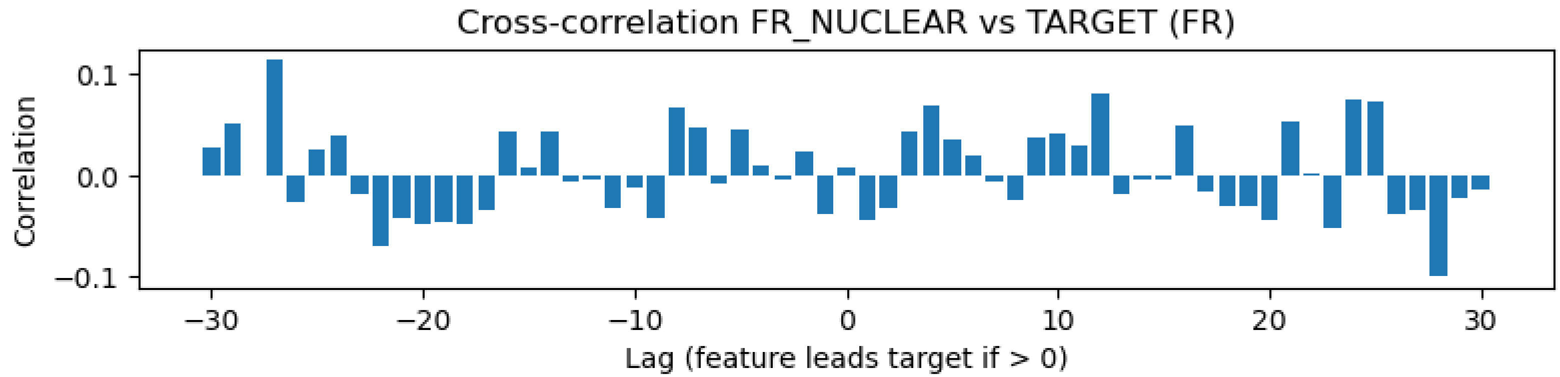

FR_GAS: time series with rolling means (FR)
FR_GAS
DAY_ID
FR_GAS
roll7
roll14
roll30
2
0
0
200
400
600
800
1000
1200

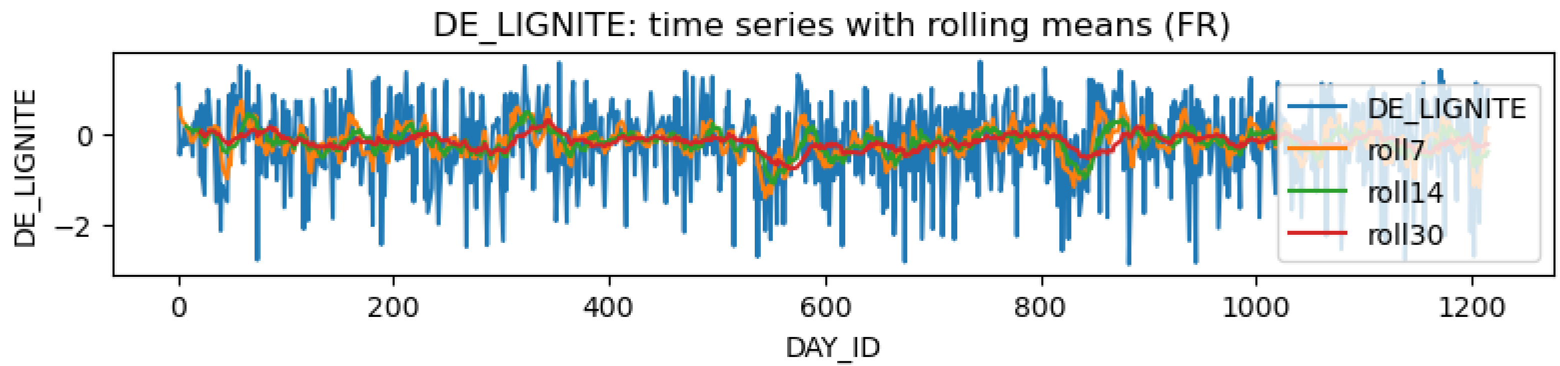

Autocorrelation of FR_GAS (FR)
ACF
Lag

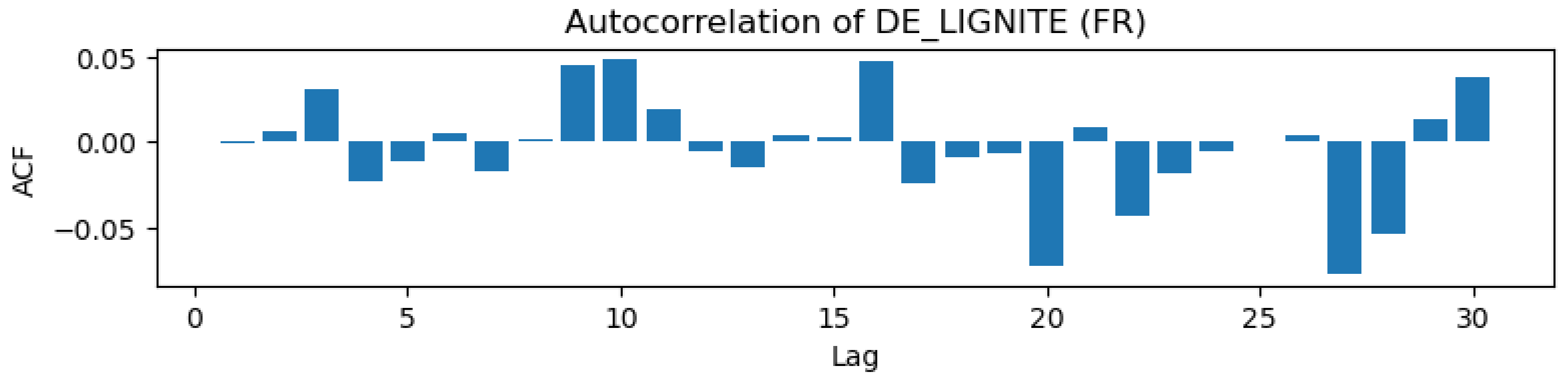

Cross-correlation FR_GAS vs TARGET (FR)
Correlation
Lag (feature leads target if > 0)

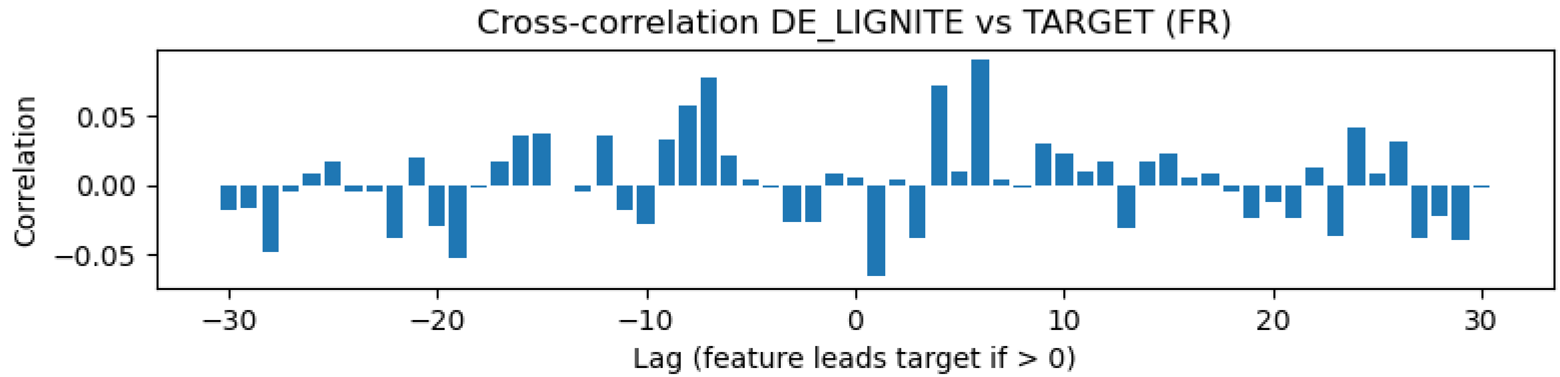

FR_NUCLEAR: time series with rolling means (FR)
FR_NUCLEAR
DAY_ID
FR_NUCLEAR
roll7
roll14
roll30

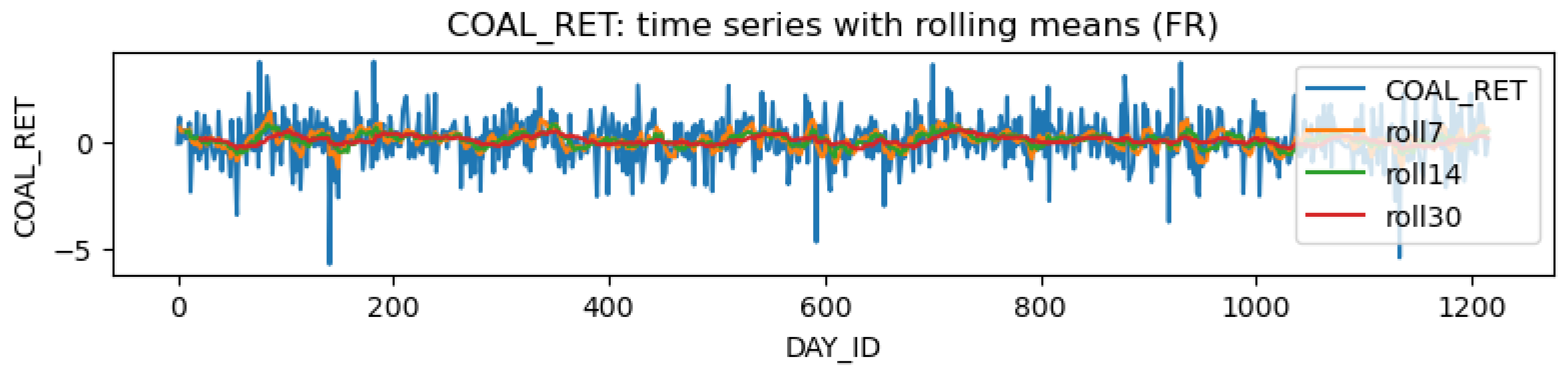

Autocorrelation of FR_NUCLEAR (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

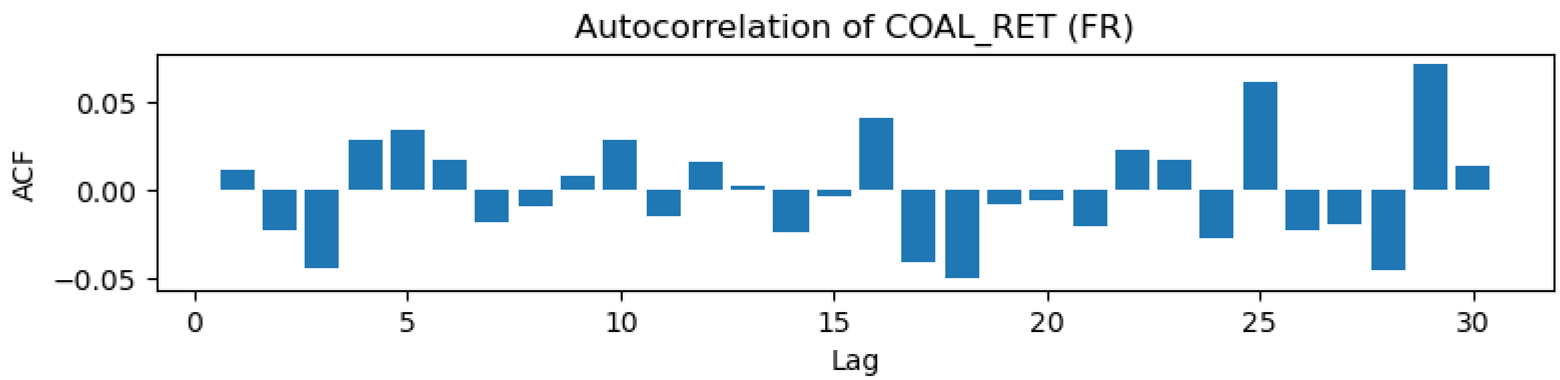

Cross-correlation FR_NUCLEAR vs TARGET (FR)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

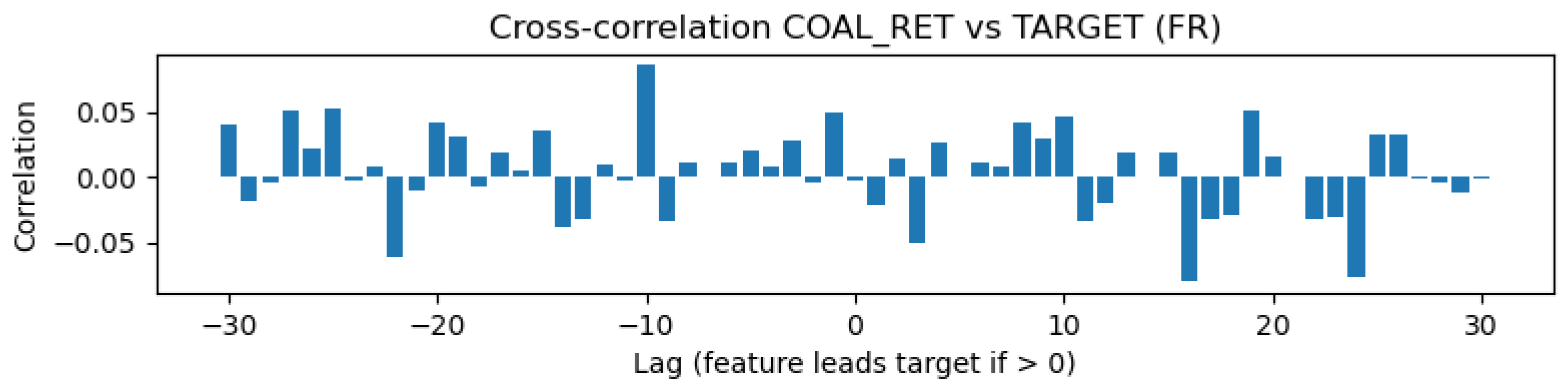

DE_LIGNITE: time series with rolling means (FR)
DE_LIGNITE
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_LIGNITE
roll7
roll14
roll30


Autocorrelation of DE_LIGNITE (FR)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

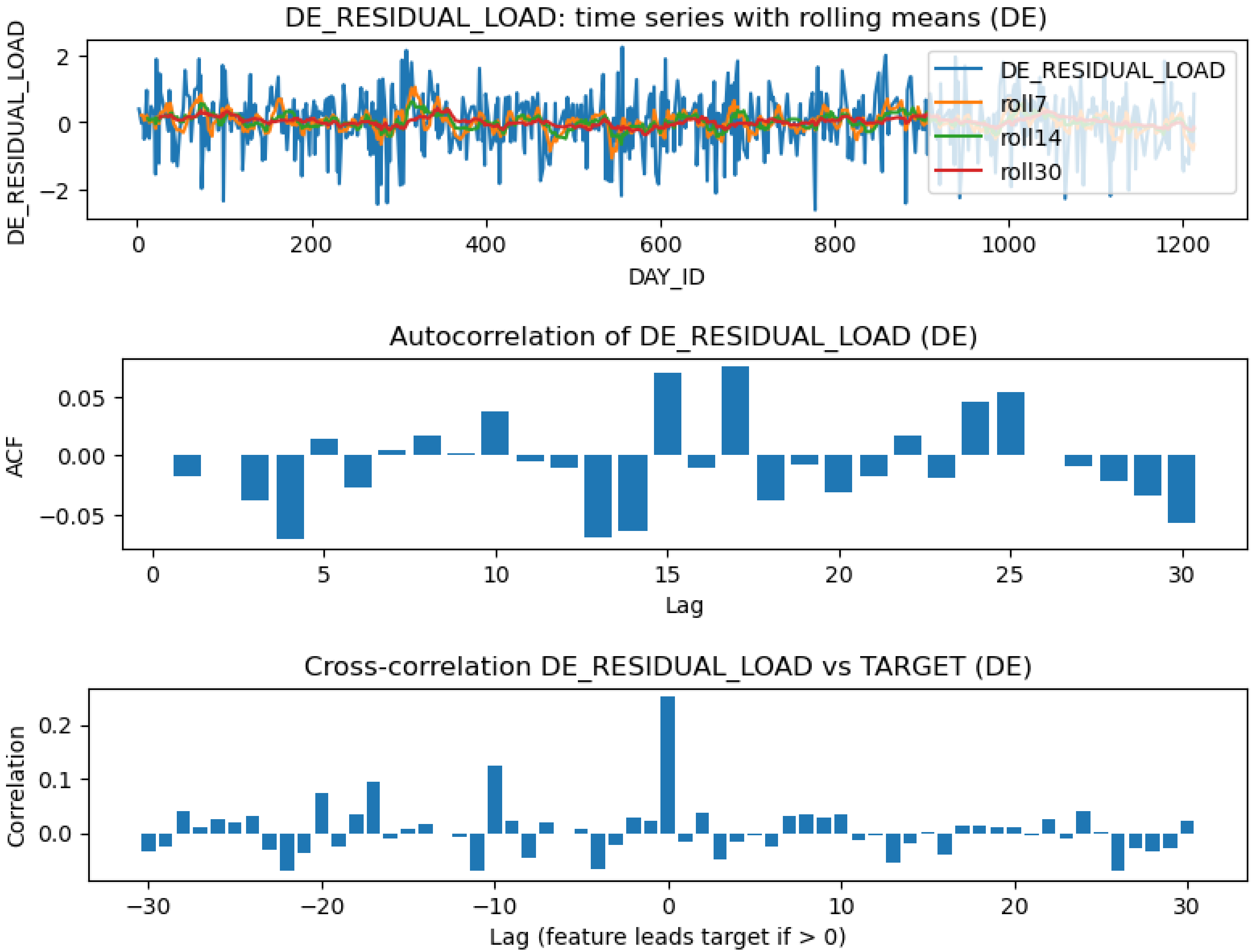

Cross-correlation DE_LIGNITE vs TARGET (FR)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

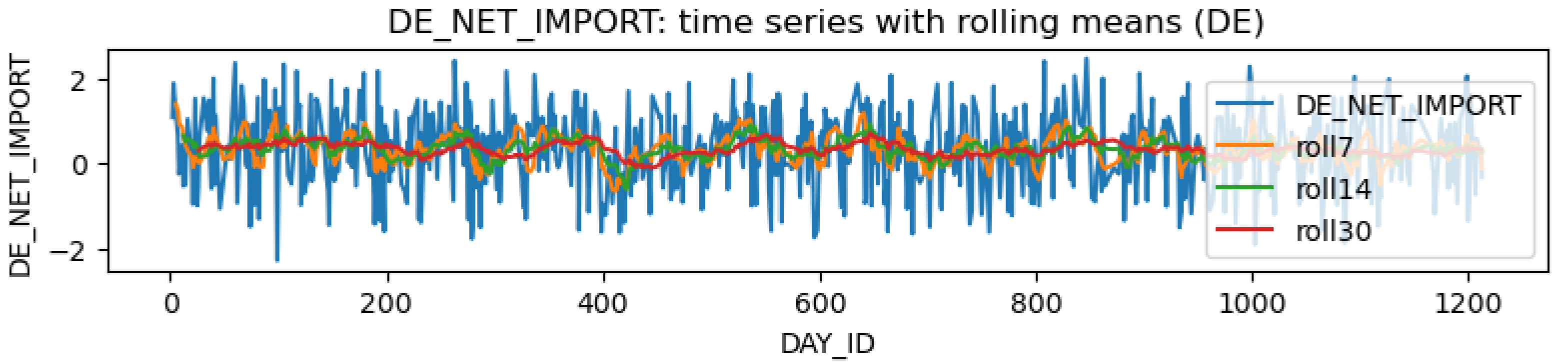

COAL_RET: time series with rolling means (FR)
COAL_RET
0
−5
0
200
400
600
800
1000
1200
DAY_ID
COAL_RET
roll7
roll14
roll30

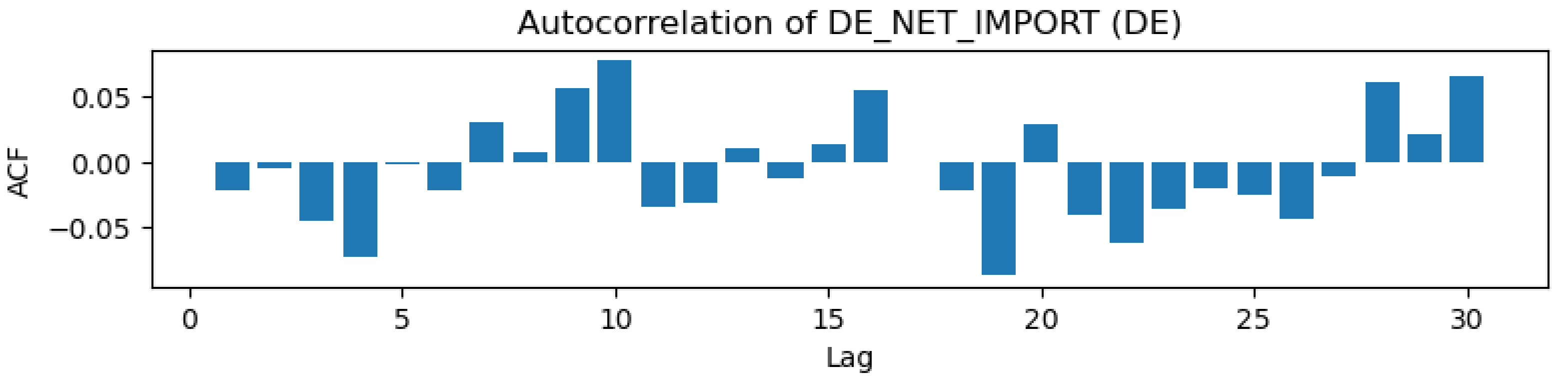


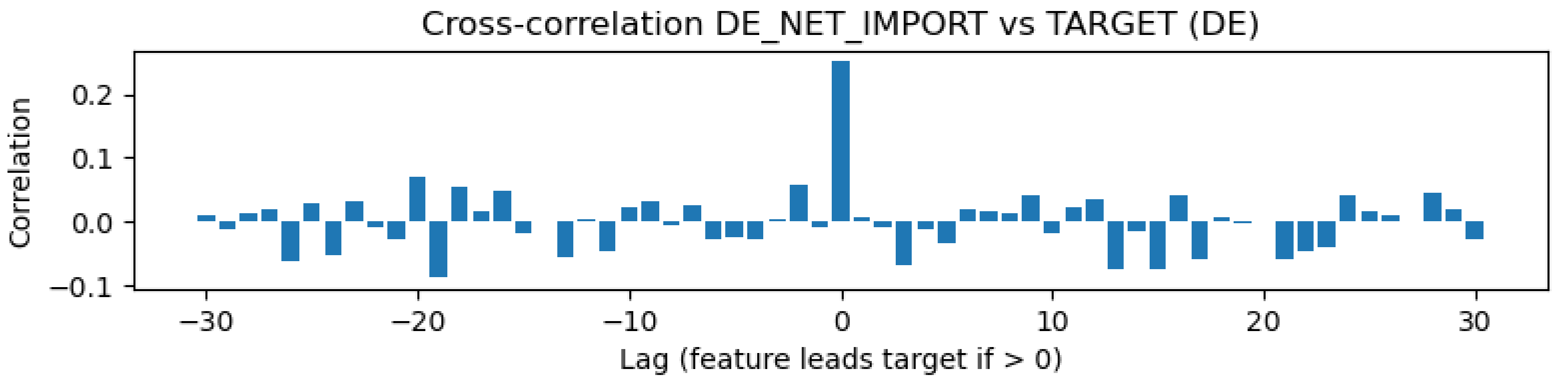


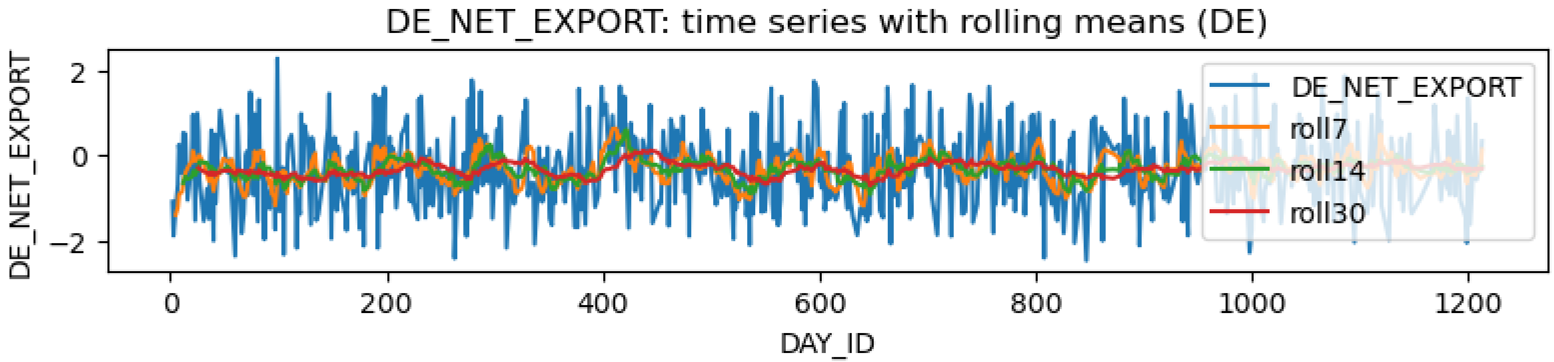

```
==============================
### COUNTRY: DE
Selected features: ['DE_RESIDUAL_LOAD', 'DE_NET_IMPORT', 'DE_NET_EXPORT', 'DE_WINDPOW', 'DE_GAS', 'FR_WINDPOW', 'DE_WIND', 'DE_HYDRO', 'DE_COAL', 'DE_LIGNITE', 'DE_FR_EXCHANGE', 'FR_DE_EXCHANGE', 'FR_WIND', 'FR_TEMP', 'DE_CONSUMPTION', 'FR_GAS', 'DE_TEMP', 'FR_RAIN', 'FR_HYDRO', 'FR_RESIDUAL_LOAD', 'FR_SOLAR', 'GAS_RET', 'COAL_RET', 'FR_CONSUMPTION', 'FR_NUCLEAR', 'DE_SOLAR', 'DE_NUCLEAR', 'DE_RAIN', 'FR_COAL', 'CARBON_RET', 'FR_NET_EXPORT', 'FR_NET_IMPORT']
```

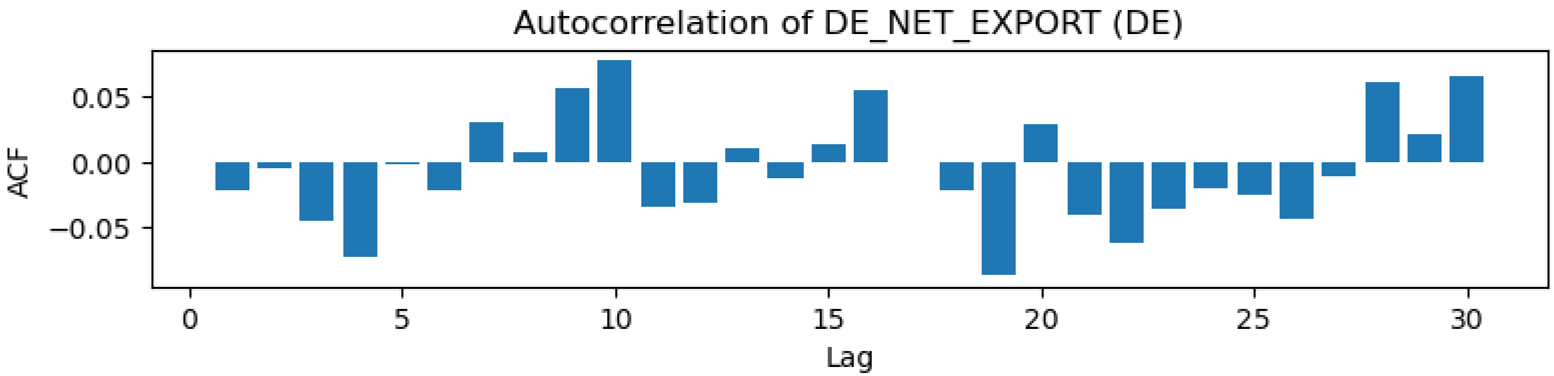

DE_RESIDUAL_LOAD: time series with rolling means (DE)
DE_RESIDUAL_LOAD
roll7
roll14
roll30
DAY_ID
Autocorrelation of DE_RESIDUAL_LOAD (DE)
ACF
Lag
Cross-correlation DE_RESIDUAL_LOAD vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

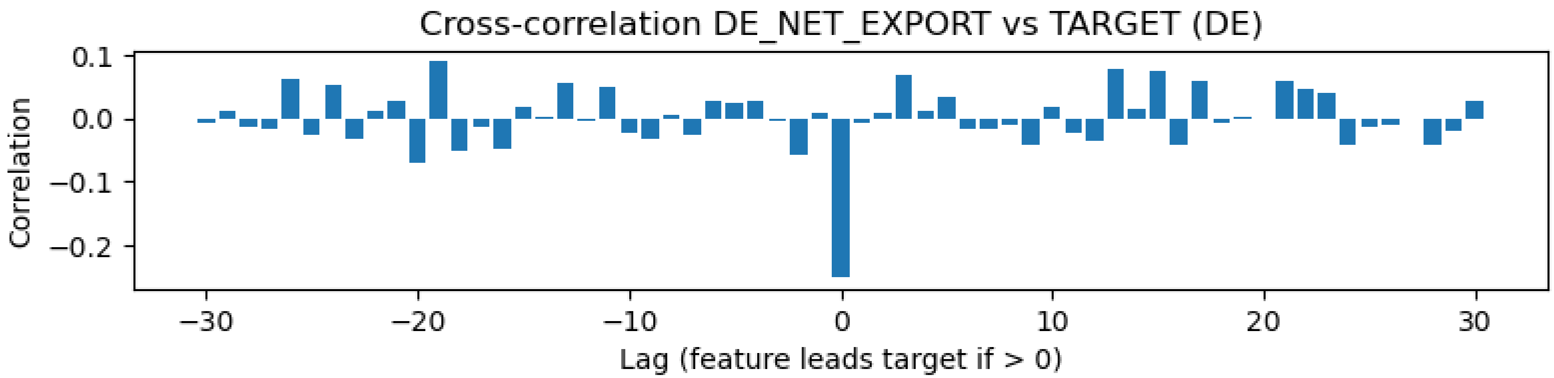

DE_NET_IMPORT: time series with rolling means (DE)
DE_NET_IMPORT
roll7
roll14
roll30
DAY_ID
DE_NET_IMPORT

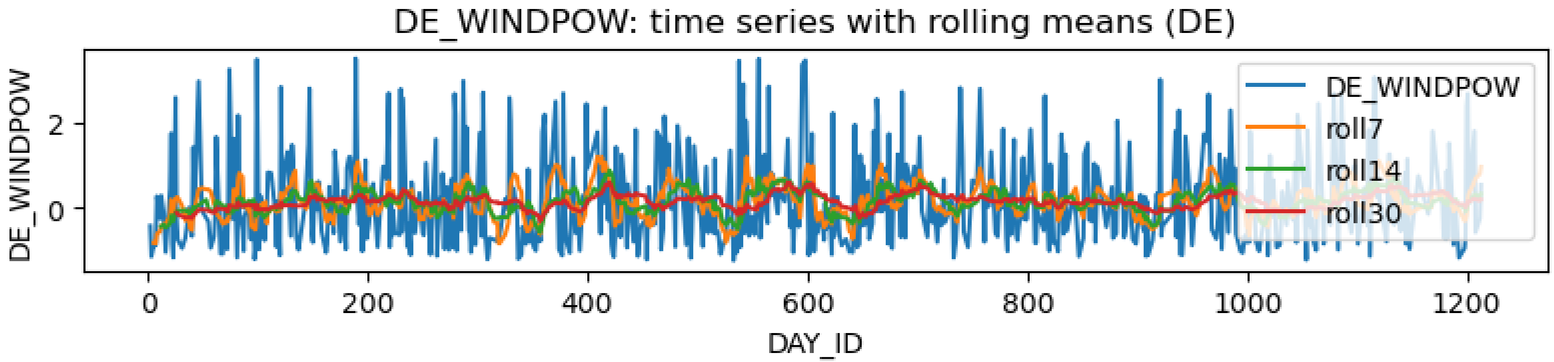

Autocorrelation of DE_NET_IMPORT (DE)
ACF
Lag

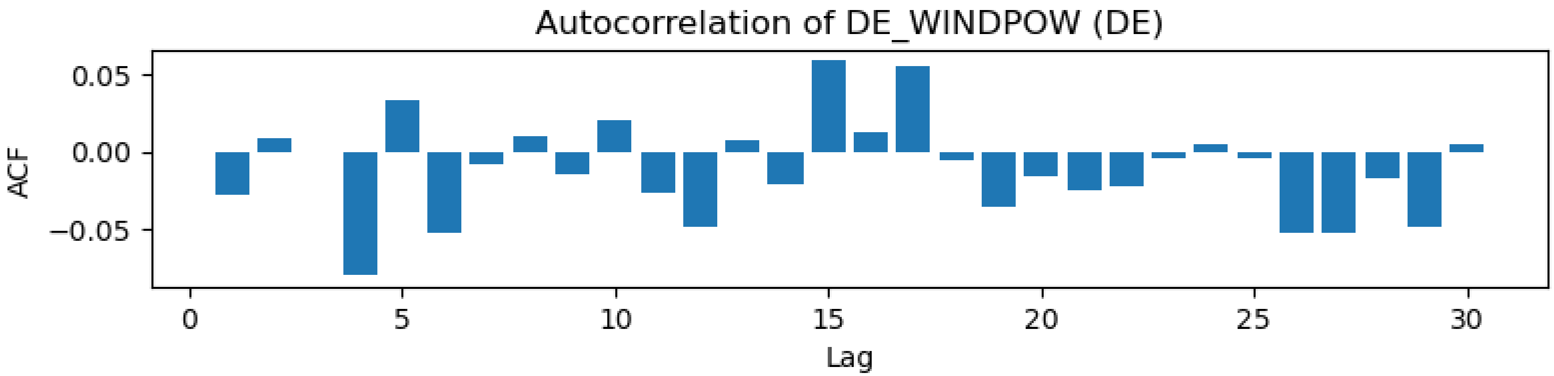

Cross-correlation DE_NET_IMPORT vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

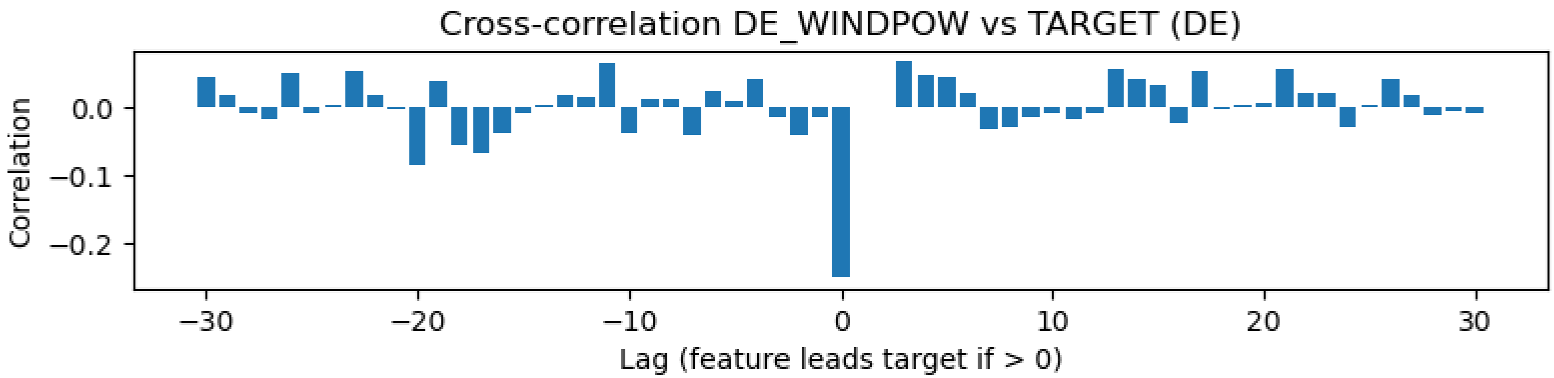

DE_NET_EXPORT: time series with rolling means (DE)
DE_NET_EXPORT
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_NET_EXPORT
roll7
roll14
roll30

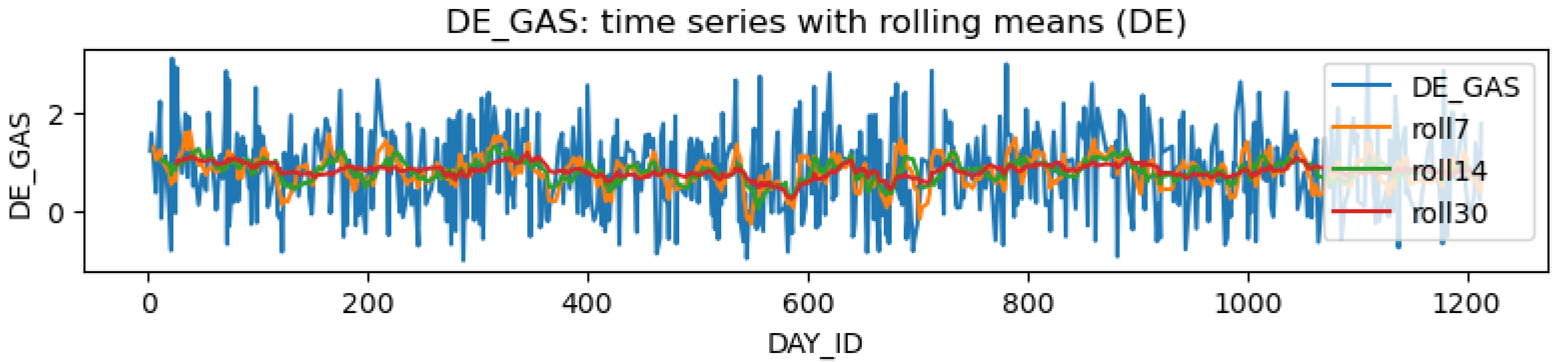

Autocorrelation of DE_NET_EXPORT (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

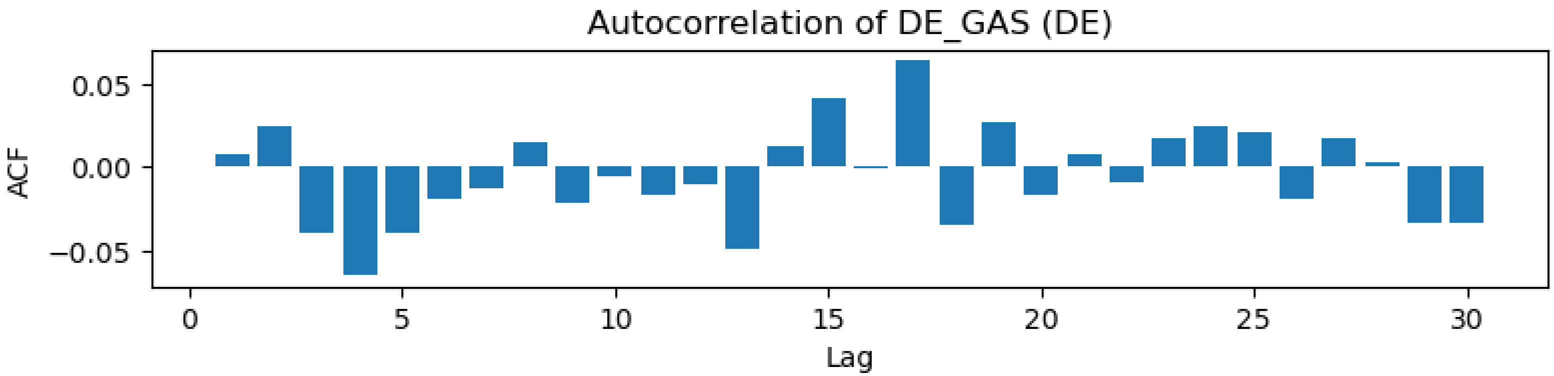

Cross-correlation DE_NET_EXPORT vs TARGET (DE)
Correlation
0.1
0.0
−0.1
−0.2
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

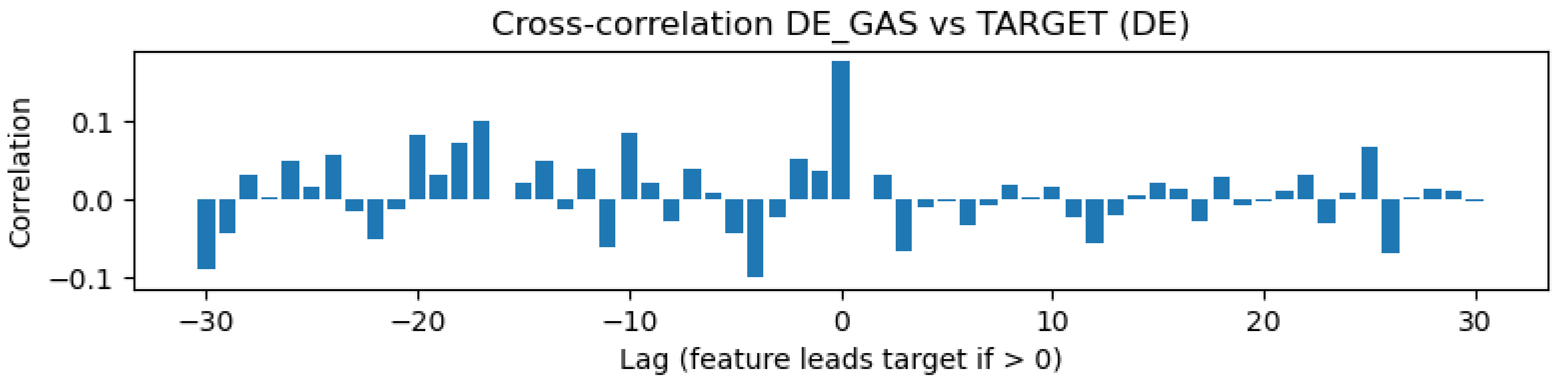

DE_WINDPOW: time series with rolling means (DE)
DE_WINDPOW
DAY_ID
DE_WINDPOW
roll7
roll14
roll30
0
2
0
200
400
600
800
1000
1200

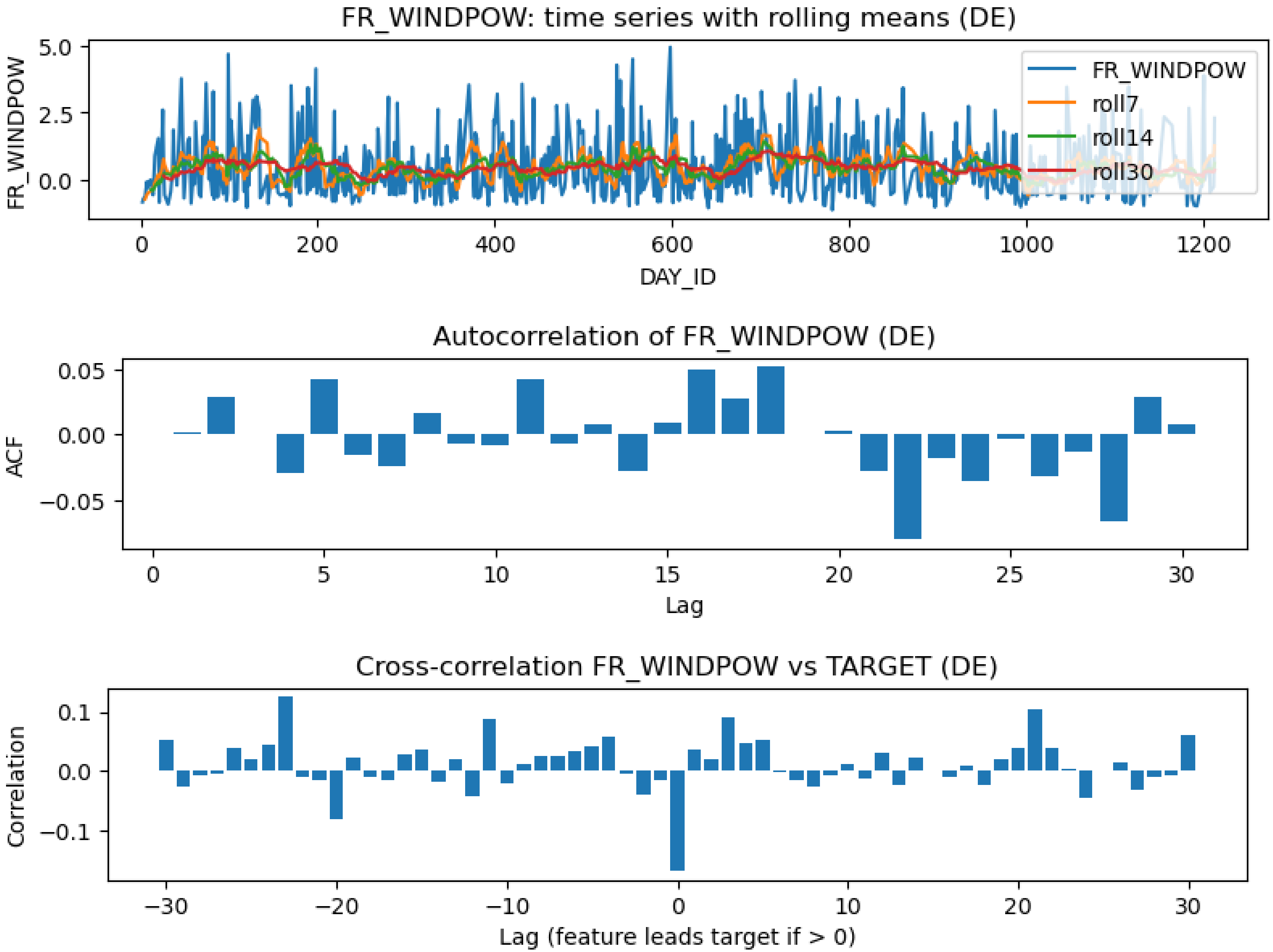

Autocorrelation of DE_WINDPOW (DE)
ACF
Lag
0.05
0.00
−0.05
0
5
10
15
20
25
30

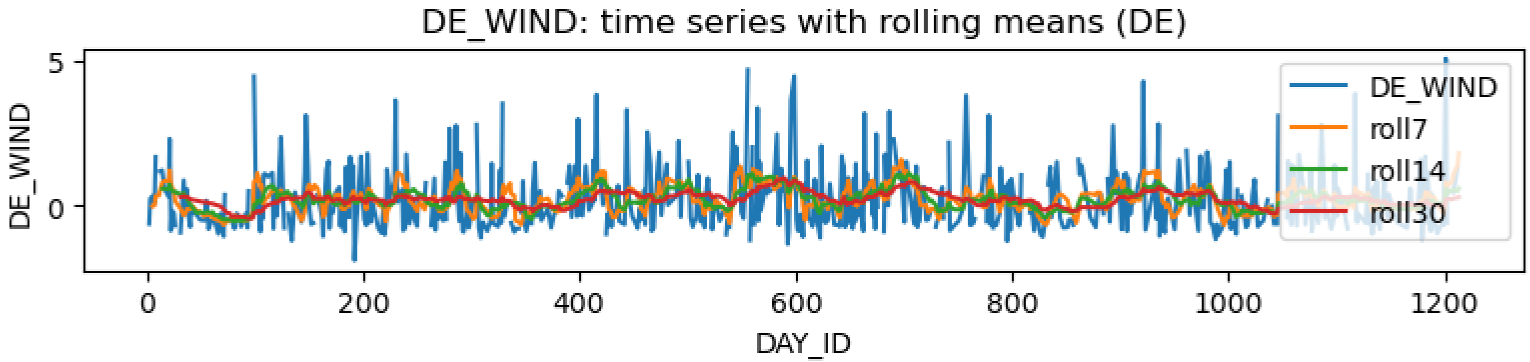

Cross-correlation DE_WINDPOW vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)
0.0
−0.1
−0.2
−30
−20
−10
0
10
20
30

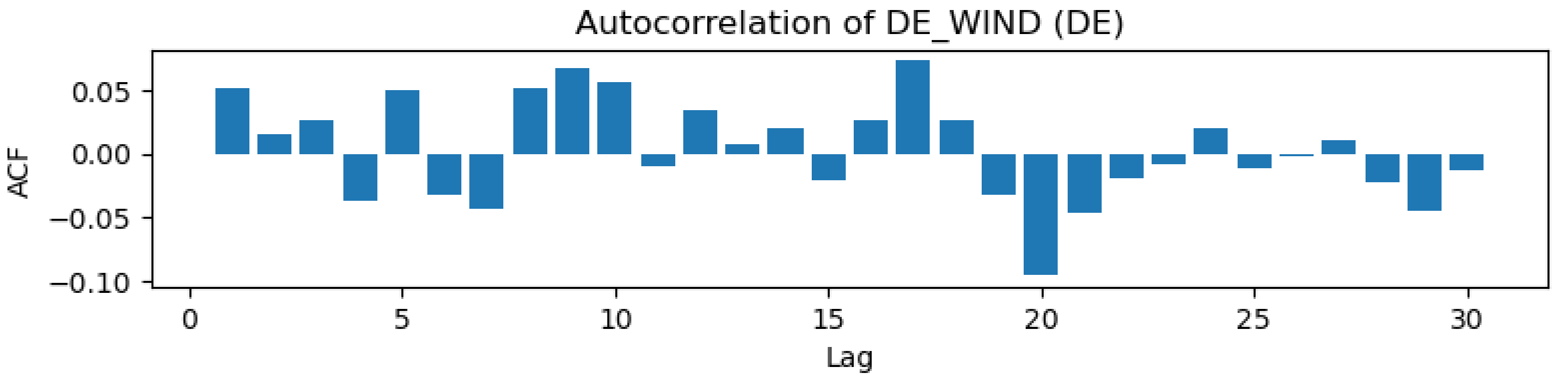

DE_GAS: time series with rolling means (DE)
DE_GAS
2
0
0
200
400
600
800
1000
1200
DAY_ID
DE_GAS
roll7
roll14
roll30

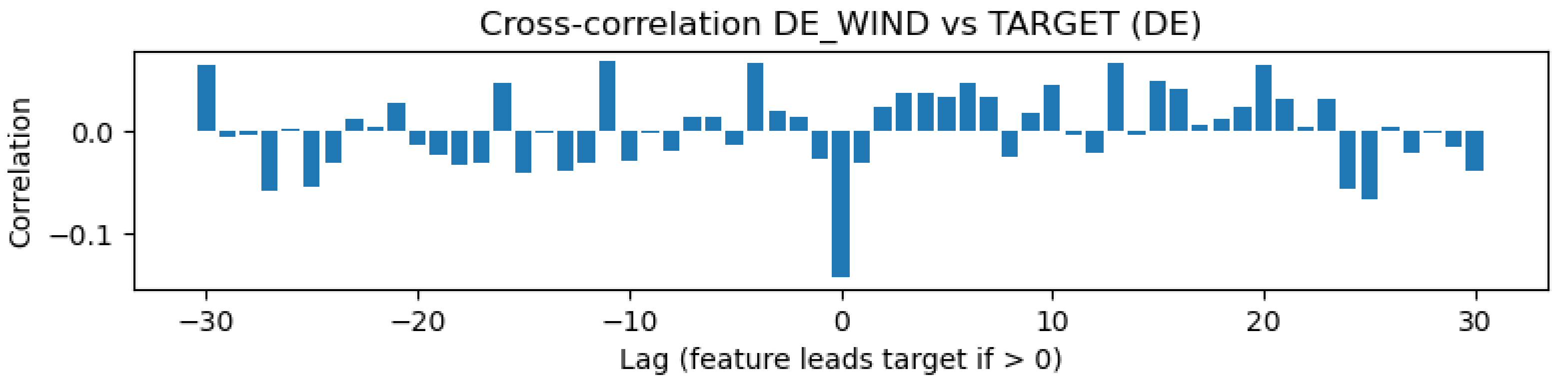

Autocorrelation of DE_GAS (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

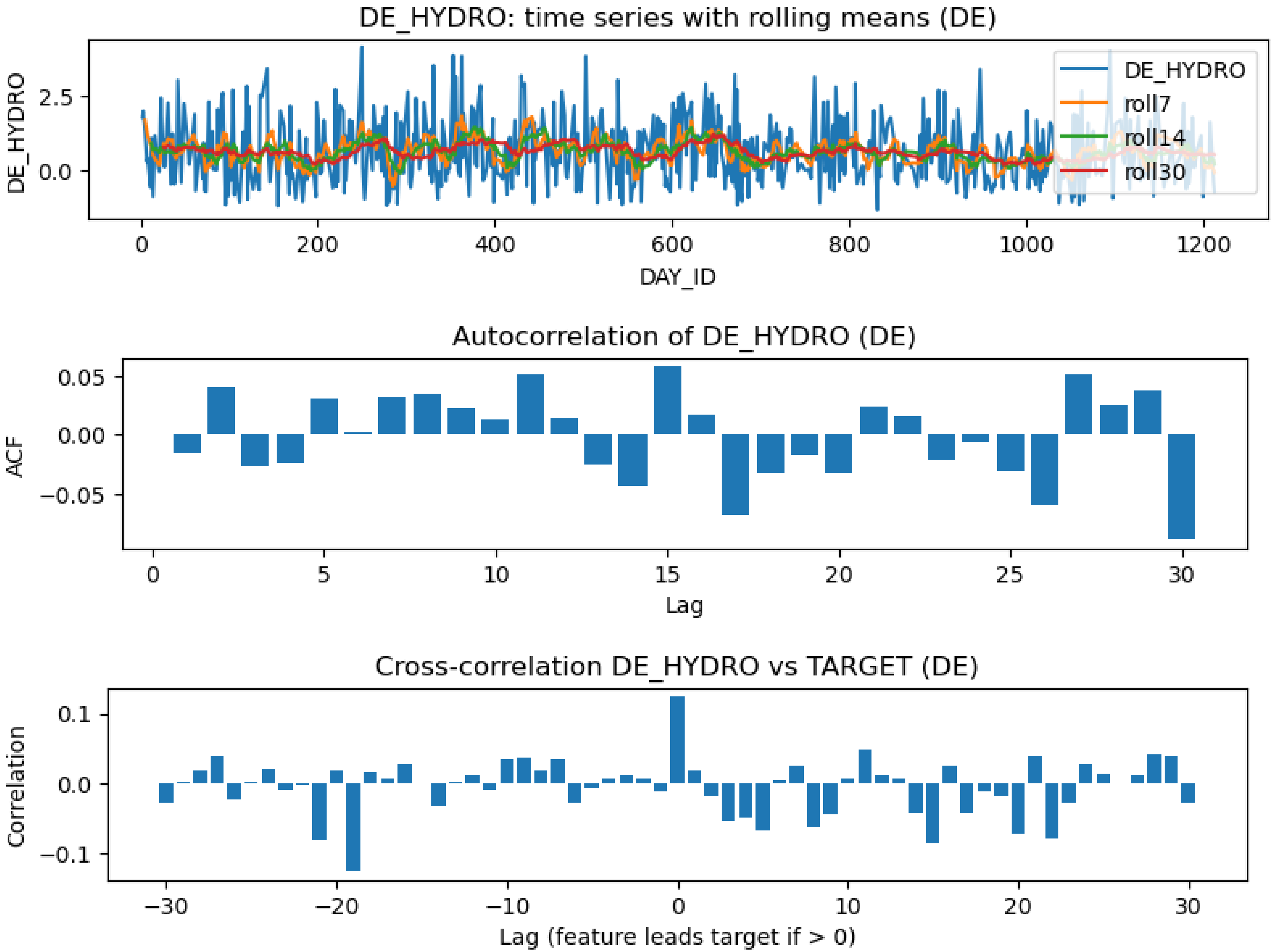

Cross-correlation DE_GAS vs TARGET (DE)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

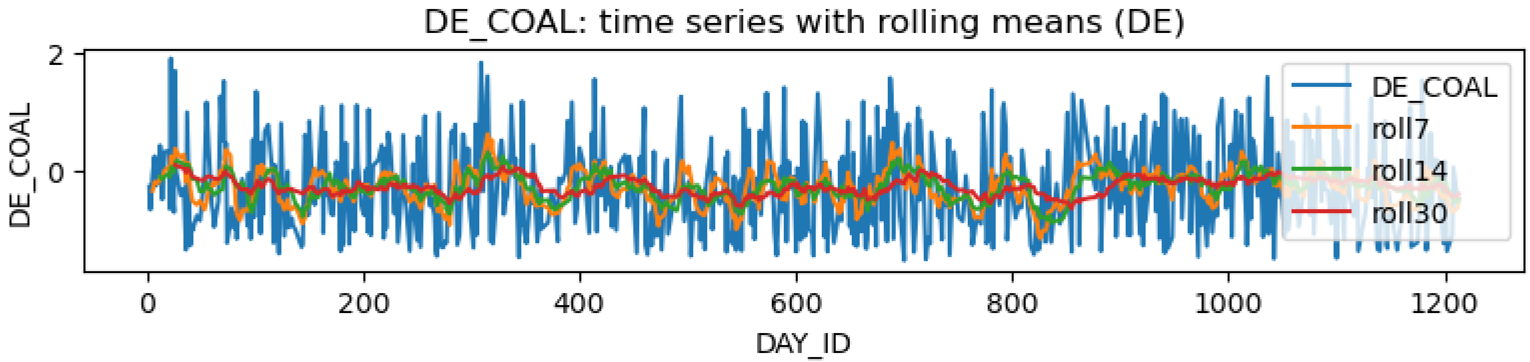

FR_WINDPOW: time series with rolling means (DE)
FR_WINDPOW
roll7
roll14
roll30
FR_WINDPOW
DAY_ID
Autocorrelation of FR_WINDPOW (DE)
ACF
Lag
Cross-correlation FR_WINDPOW vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

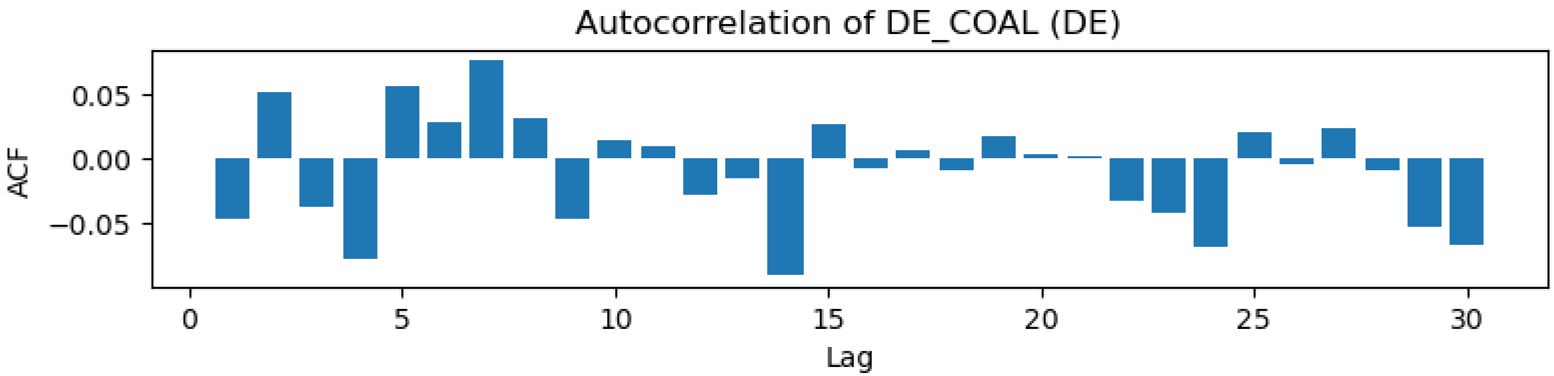

DE_WIND: time series with rolling means (DE)
DE_WIND
roll7
roll14
roll30
DAY_ID
0
200
400
600
800
1000
1200
5
0

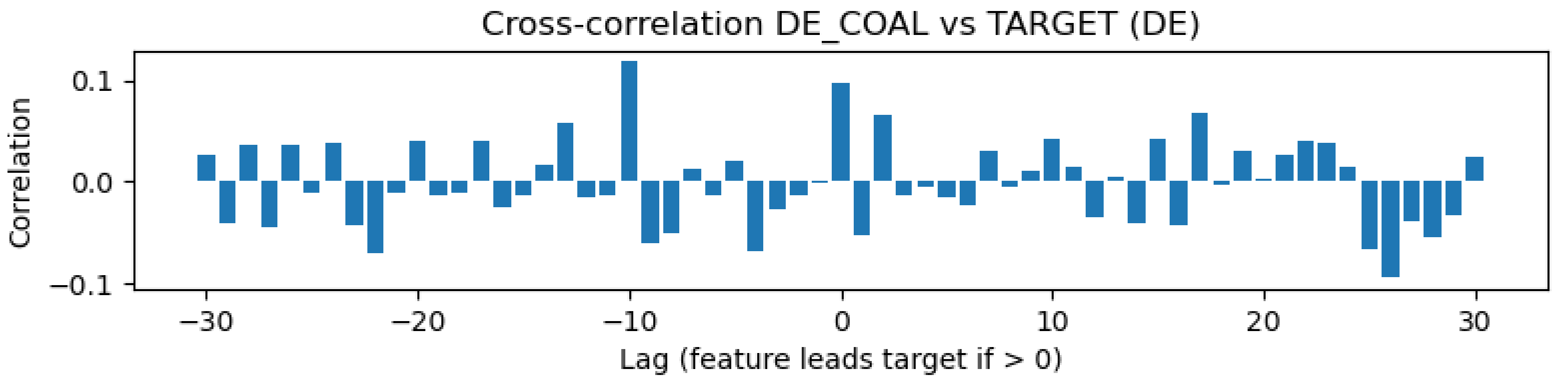

Autocorrelation of DE_WIND (DE)
ACF
Lag
0.05
0.00
−0.05
−0.10
0
5
10
15
20
25
30

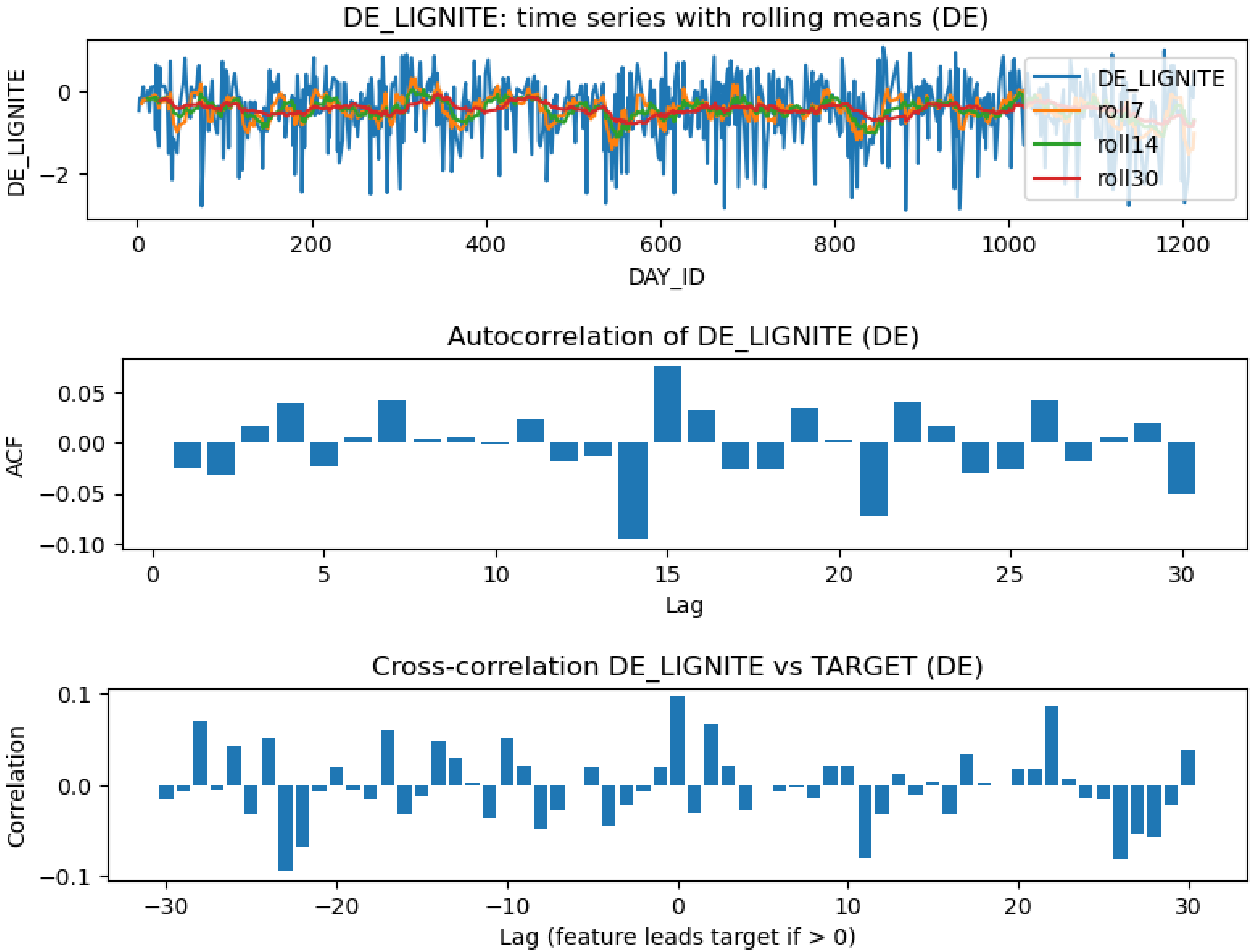

Cross-correlation DE_WIND vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)
0.0
−0.1
−30
−20
−10
0
10
20
30

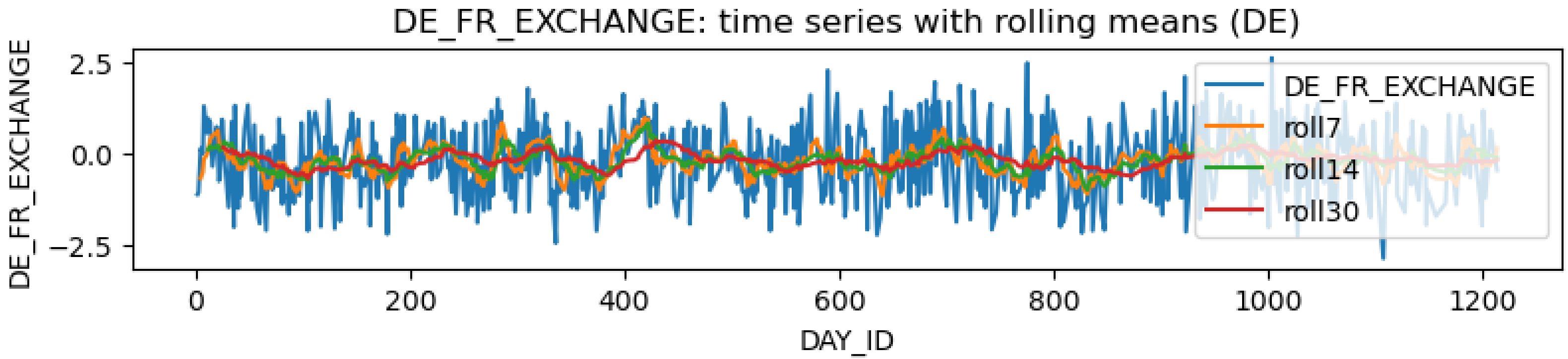

DE_HYDRO: time series with rolling means (DE)
DE_HYDRO
roll7
roll14
roll30
DE_HYDRO
DAY_ID
Autocorrelation of DE_HYDRO (DE)
ACF
Lag
Cross-correlation DE_HYDRO vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

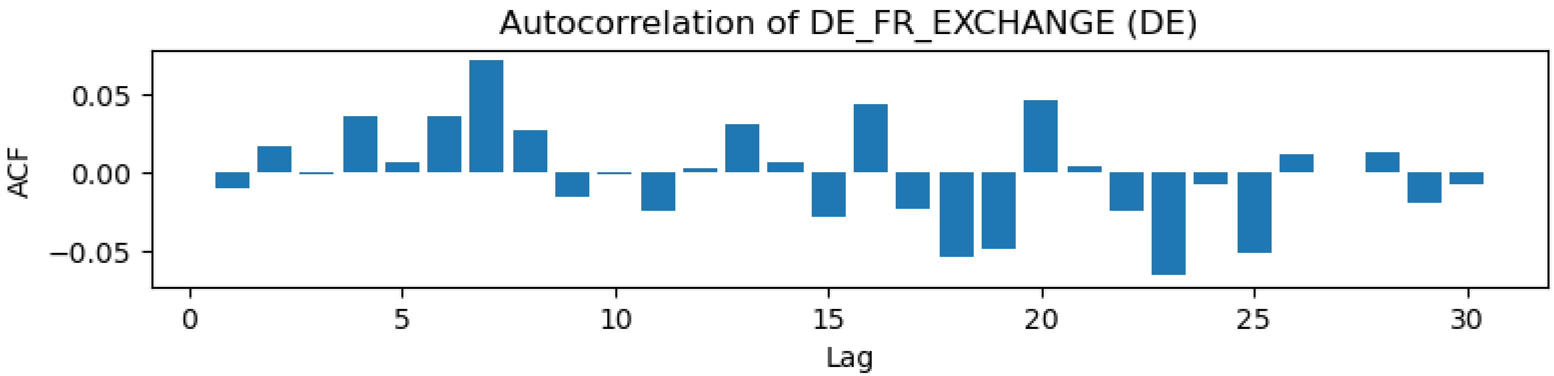

DE_COAL: time series with rolling means (DE)
DE_COAL
roll7
roll14
roll30
DE_COAL
DAY_ID
0
200
400
600
800
1000
1200
2
0

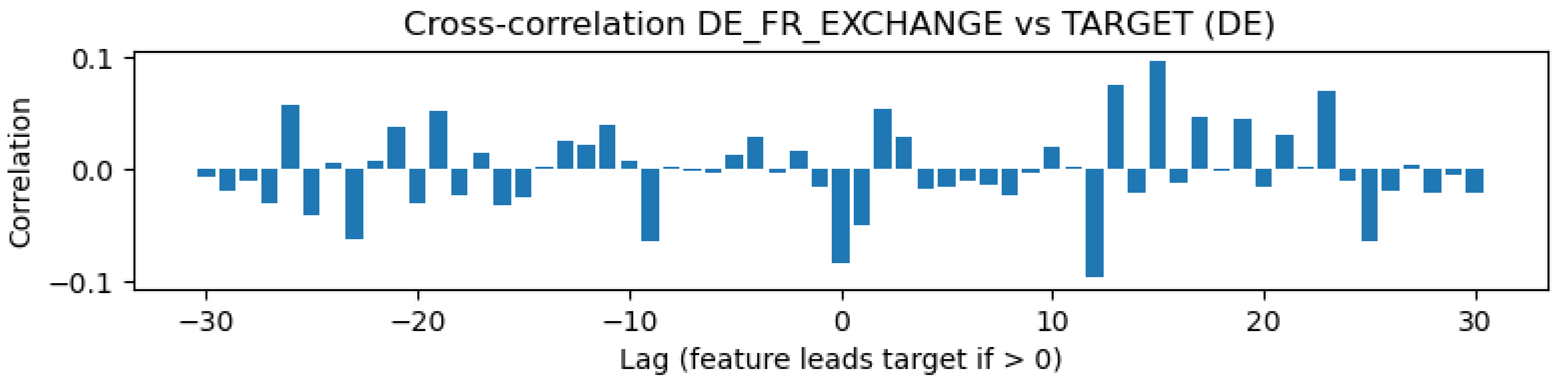

Autocorrelation of DE_COAL (DE)
ACF
Lag
0.05
0.00
−0.05
0
5
10
15
20
25
30

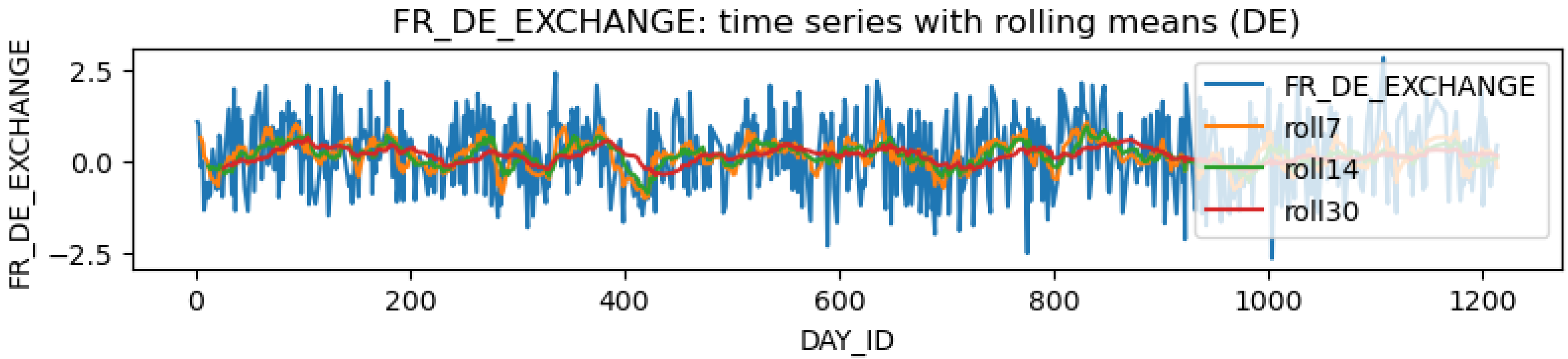

Cross-correlation DE_COAL vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30

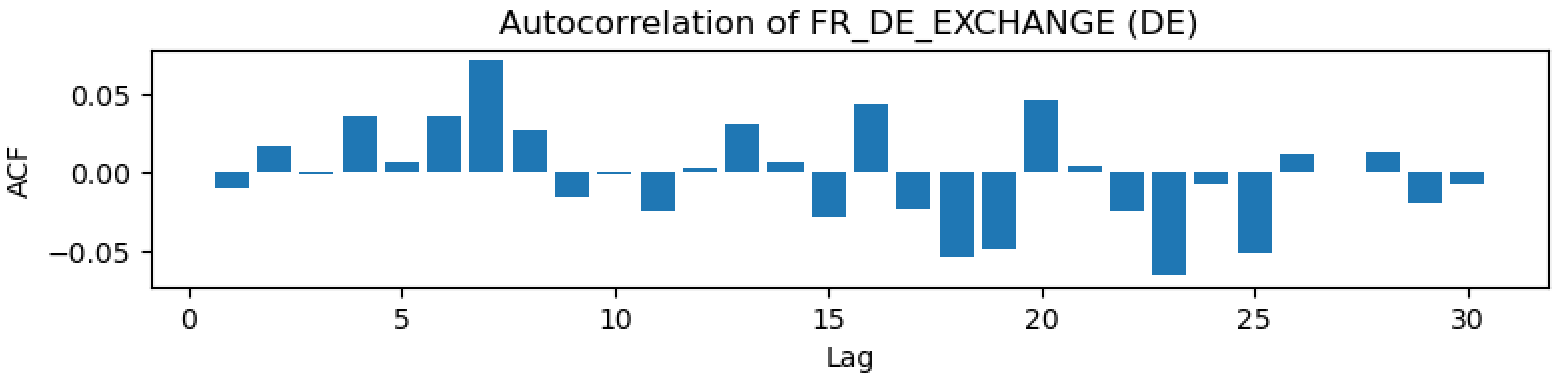

DE_LIGNITE: time series with rolling means (DE)
DE_LIGNITE
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_LIGNITE
roll7
roll14
roll30
Autocorrelation of DE_LIGNITE (DE)
ACF
0.05
0.00
−0.05
−0.10
0
5
10
15
20
25
30
Lag
Cross-correlation DE_LIGNITE vs TARGET (DE)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

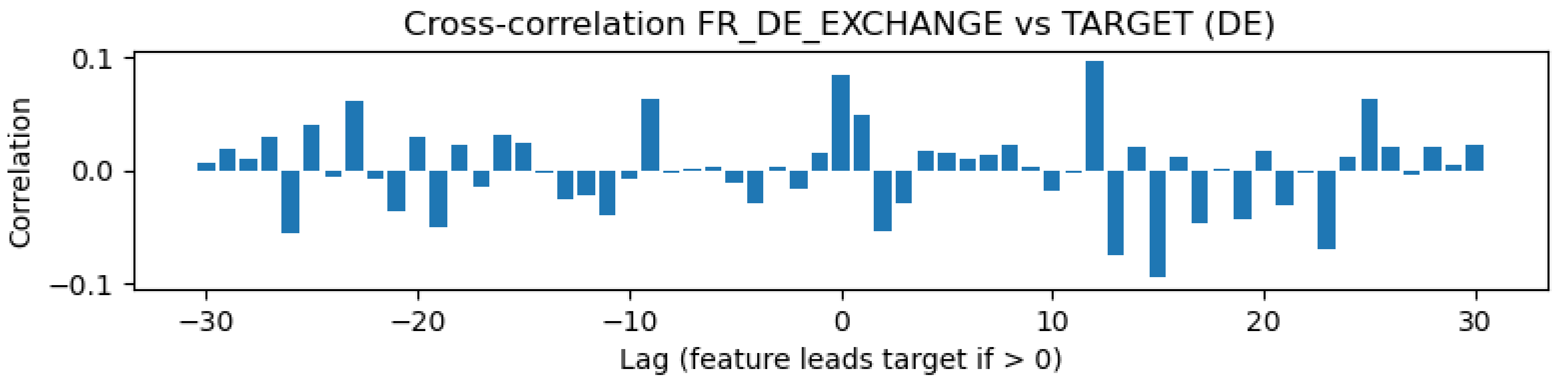

DE_FR_EXCHANGE: time series with rolling means (DE)
DE_FR_EXCHANGE
2.5
0.0
−2.5
0
200
400
600
800
1000
1200
DAY_ID
DE_FR_EXCHANGE
roll7
roll14
roll30

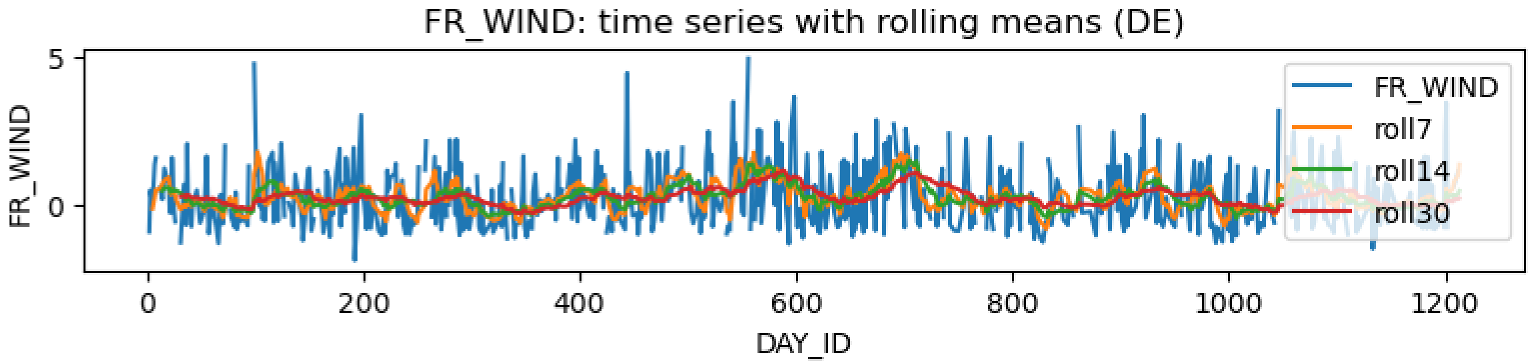

Autocorrelation of DE_FR_EXCHANGE (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

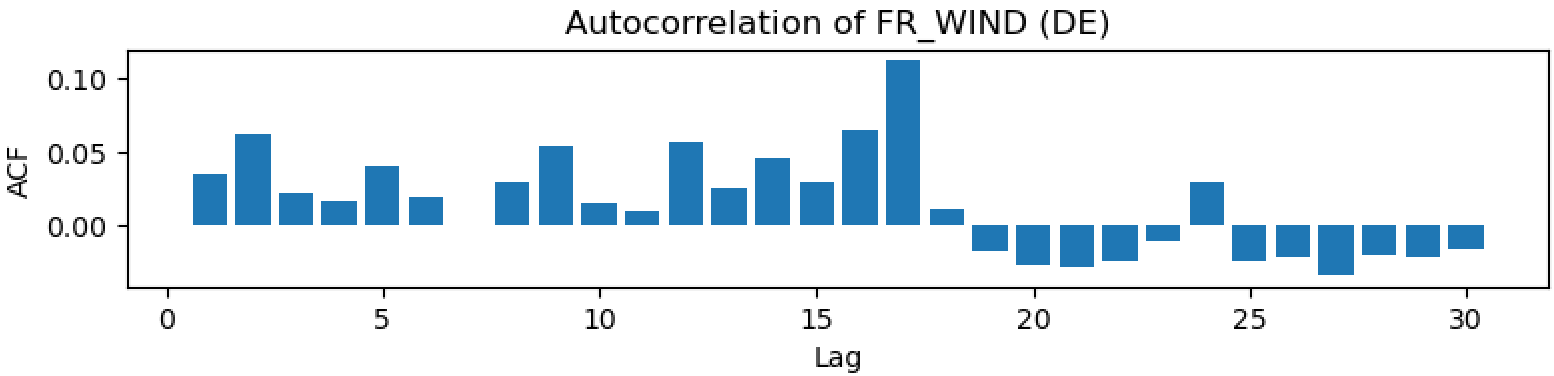

Cross-correlation DE_FR_EXCHANGE vs TARGET (DE)
Correlation
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

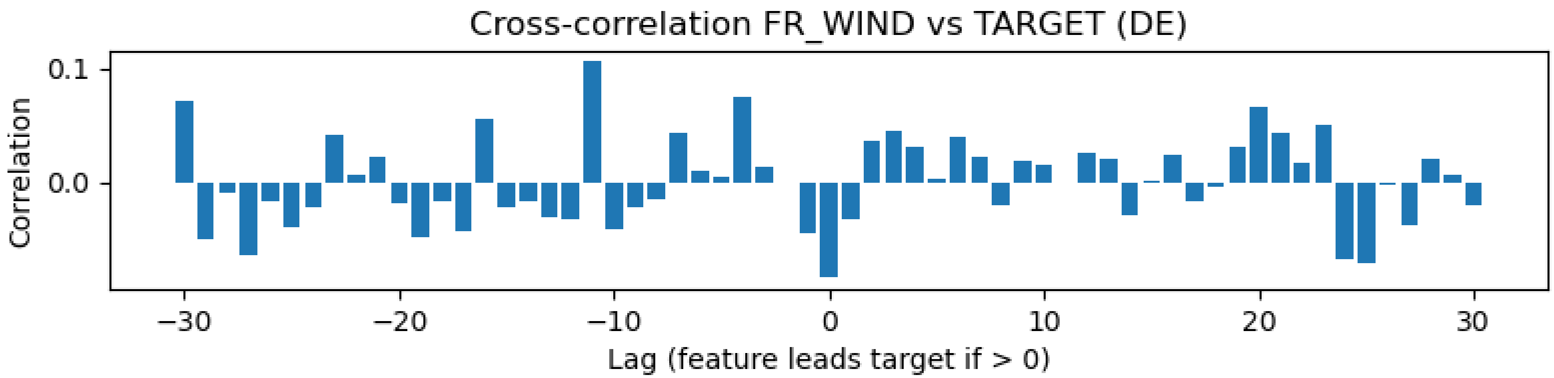

FR_DE_EXCHANGE: time series with rolling means (DE)
FR_DE_EXCHANGE
roll7
roll14
roll30
DAY_ID
2.5
0.0
−2.5
0
200
400
600
800
1000
1200

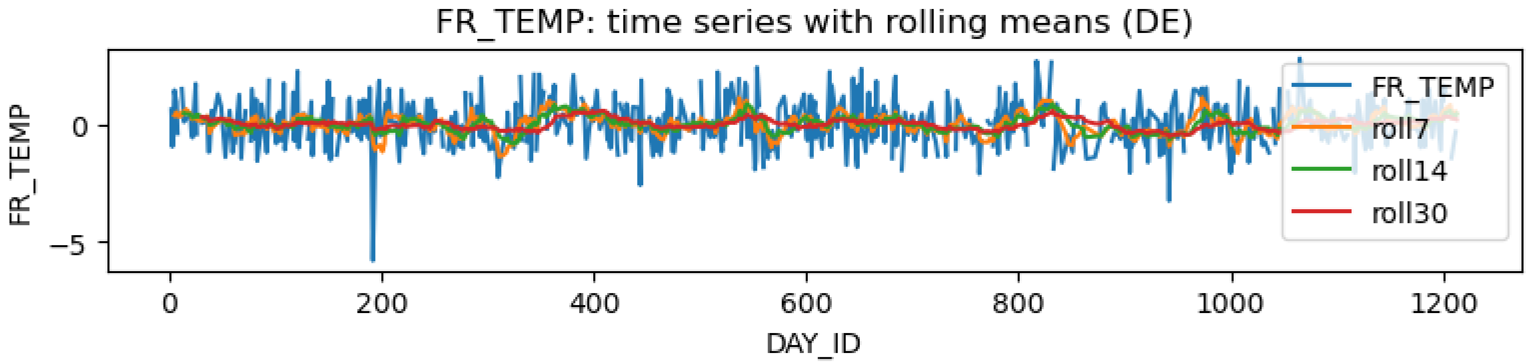

Autocorrelation of FR_DE_EXCHANGE (DE)
ACF
Lag
0.05
0.00
−0.05
0
5
10
15
20
25
30

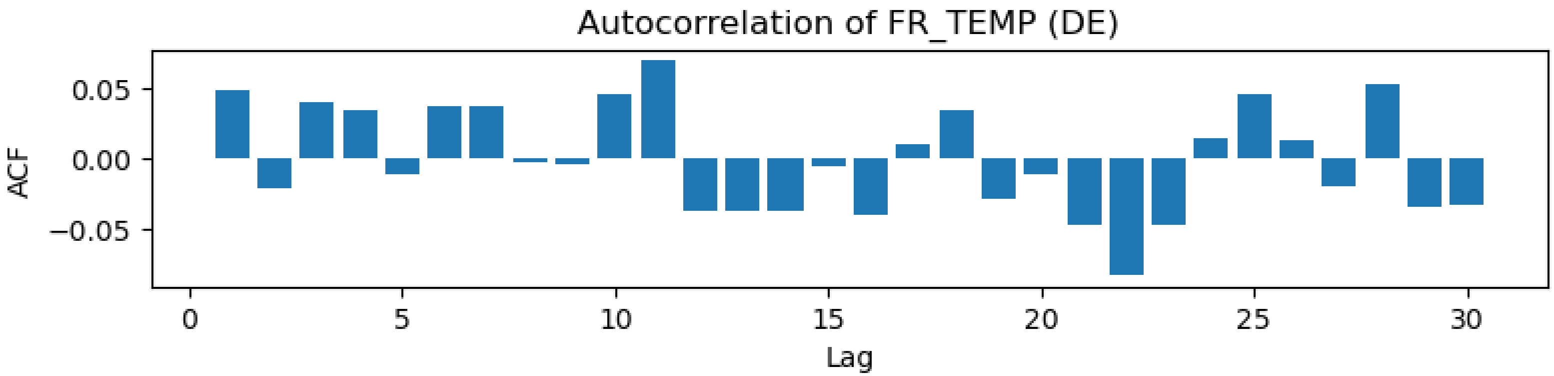

Cross-correlation FR_DE_EXCHANGE vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30

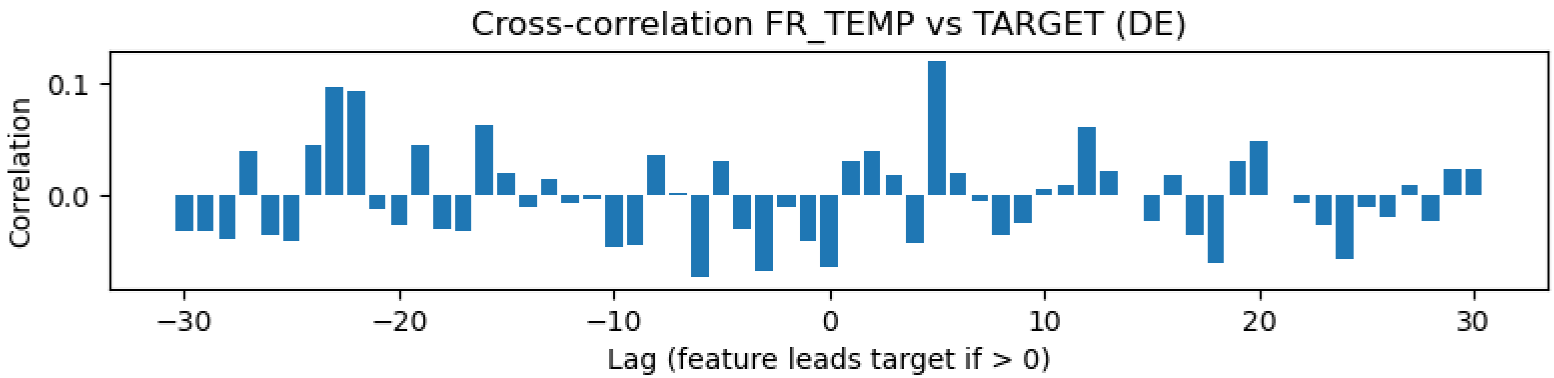

FR_WIND: time series with rolling means (DE)
FR_WIND
DAY_ID
roll7
roll14
roll30
0
5
200
400
600
800
1000
1200

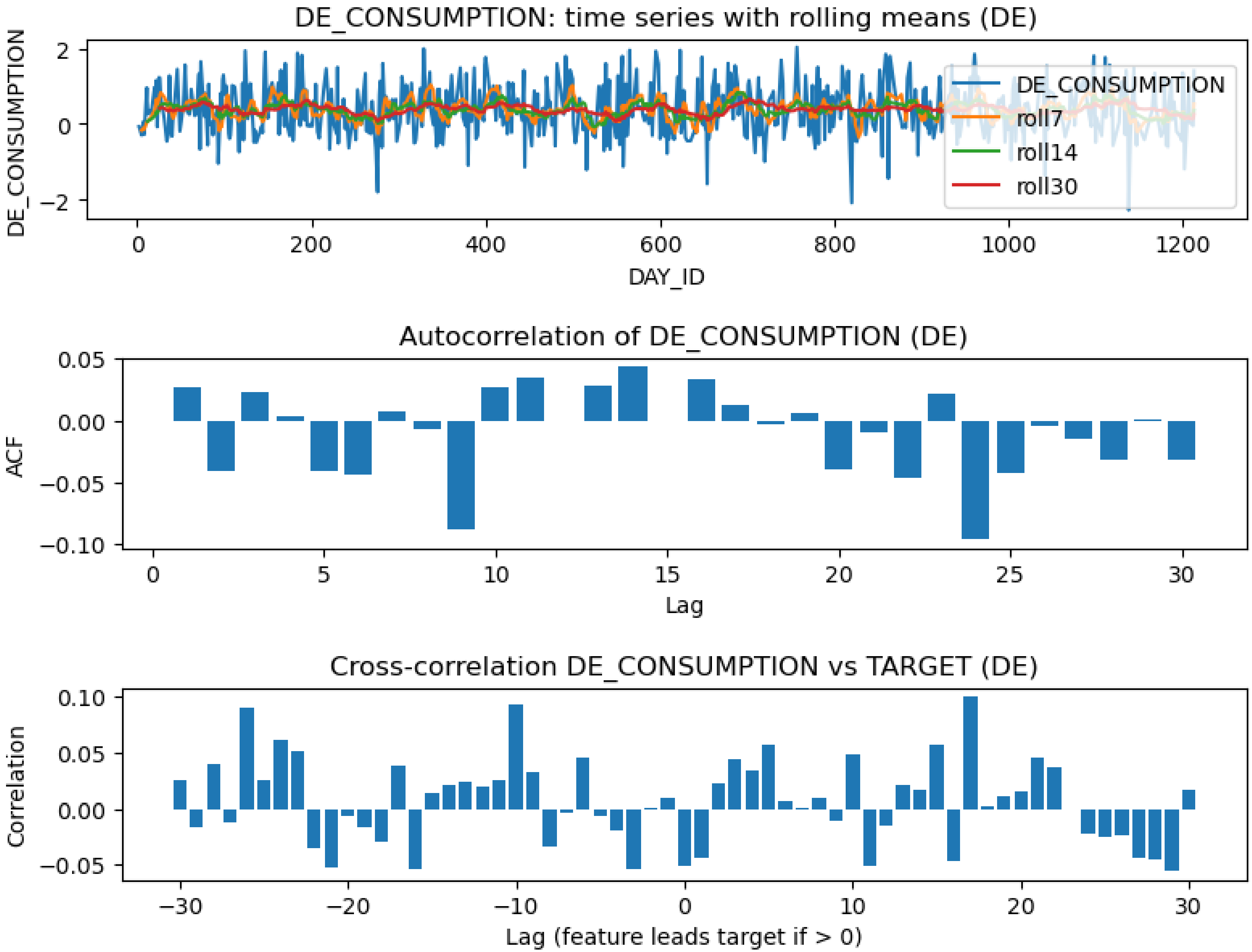

Autocorrelation of FR_WIND (DE)
ACF
Lag
0.10
0.05
0.00
0
5
10
15
20
25
30

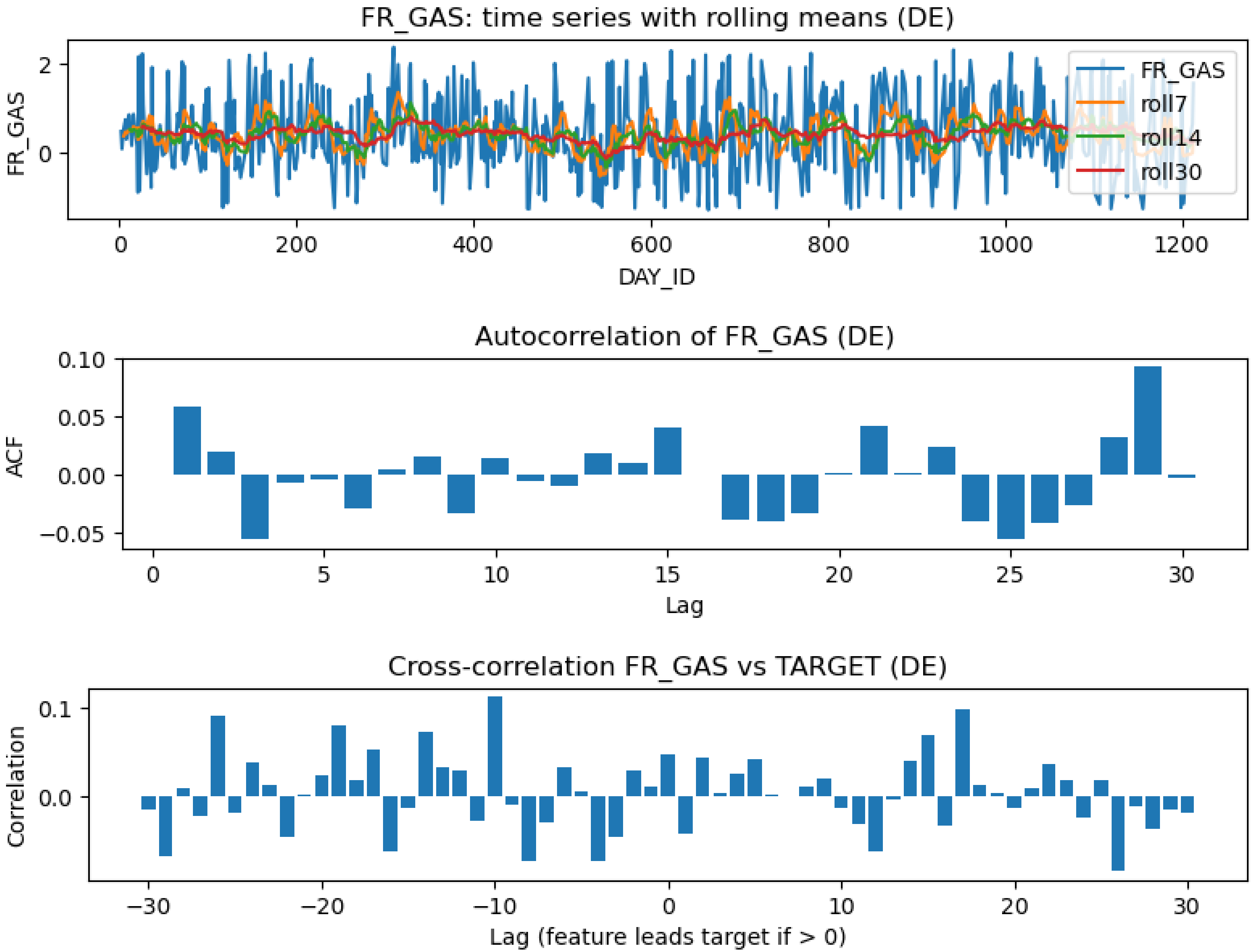

Cross-correlation FR_WIND vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)
0.1
0.0
−30
−20
−10
0
10
20
30

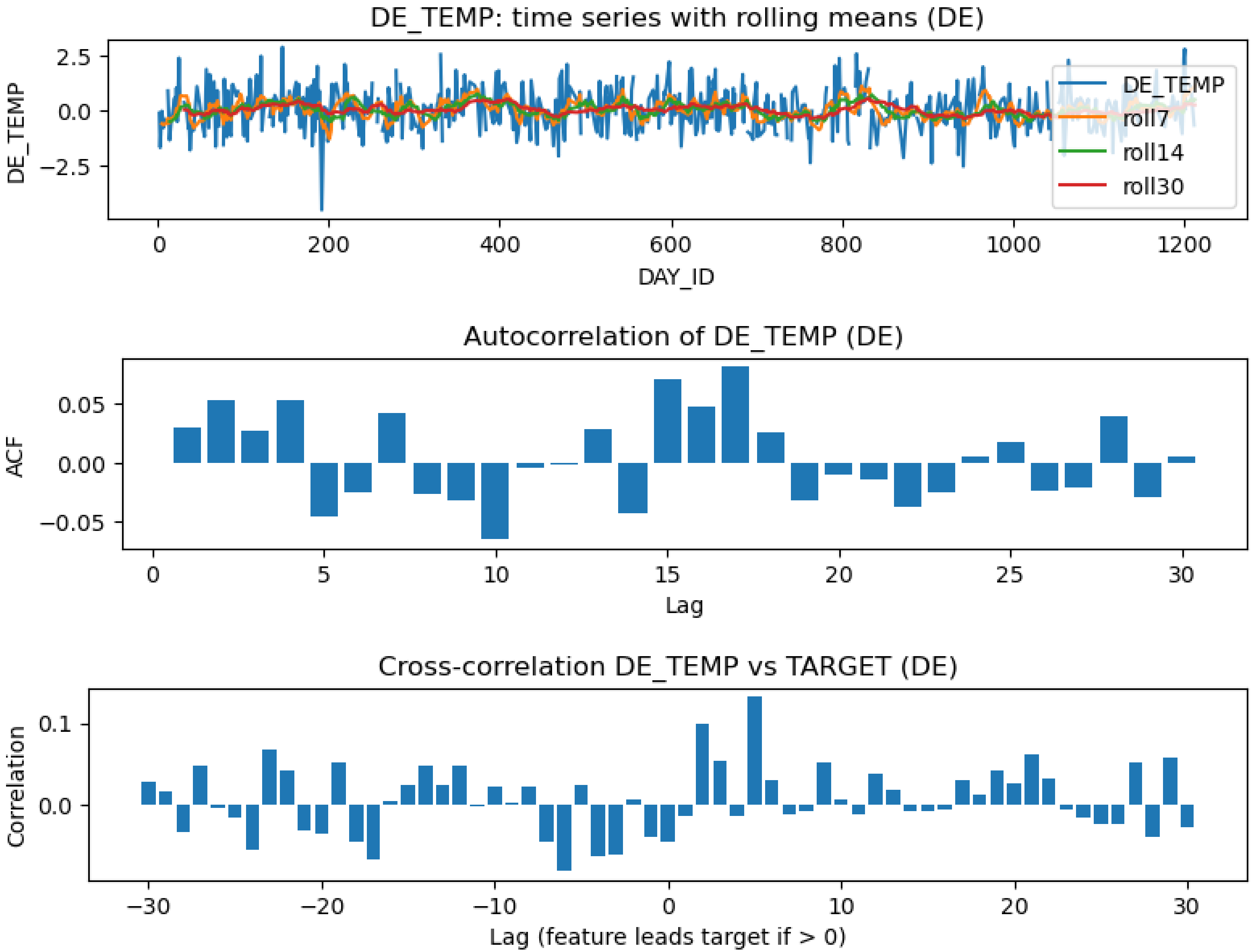

FR_TEMP: time series with rolling means (DE)
FR_TEMP
roll7
roll14
roll30
0
−5
0
200
400
600
800
1000
1200
DAY_ID

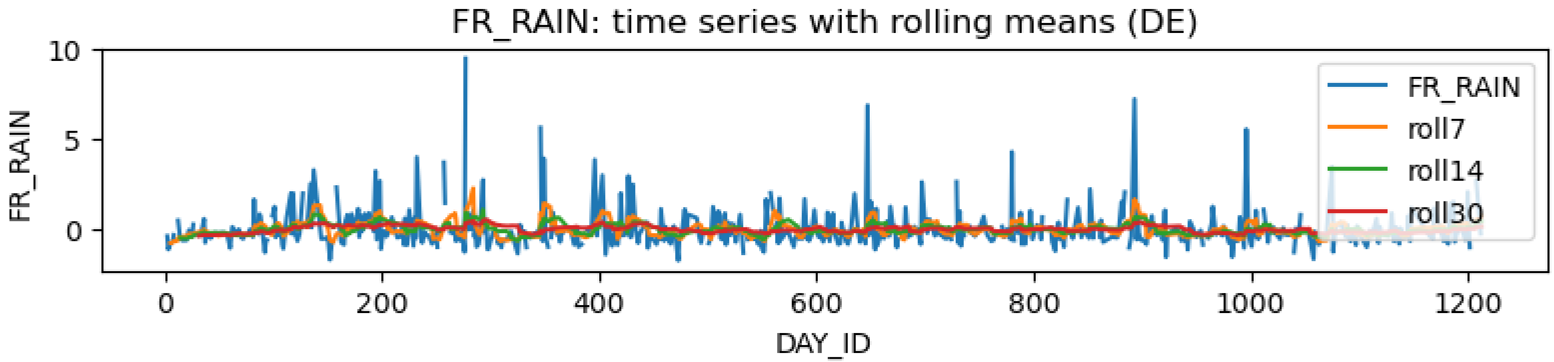

Autocorrelation of FR_TEMP (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

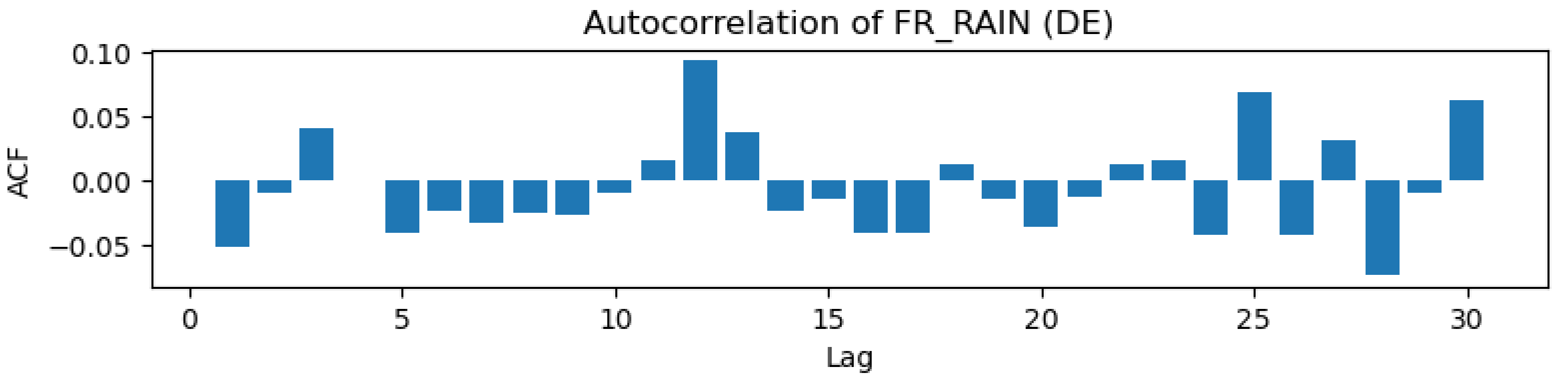

Cross-correlation FR_TEMP vs TARGET (DE)
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

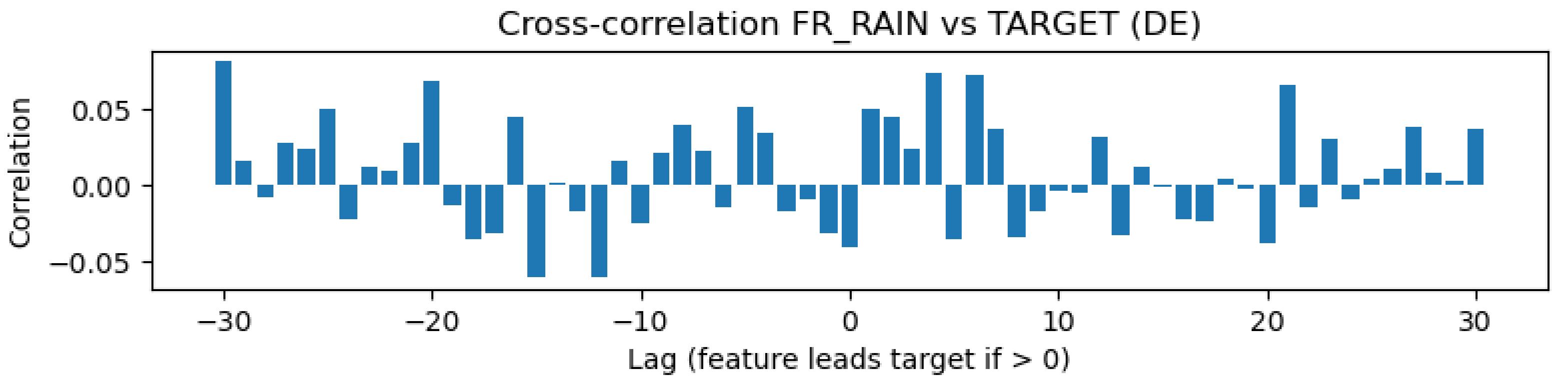

DE_CONSUMPTION: time series with rolling means (DE)
DE_CONSUMPTION
DAY_ID
DE_CONSUMPTION
roll7
roll14
roll30
Autocorrelation of DE_CONSUMPTION (DE)
ACF
Lag
Cross-correlation DE_CONSUMPTION vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

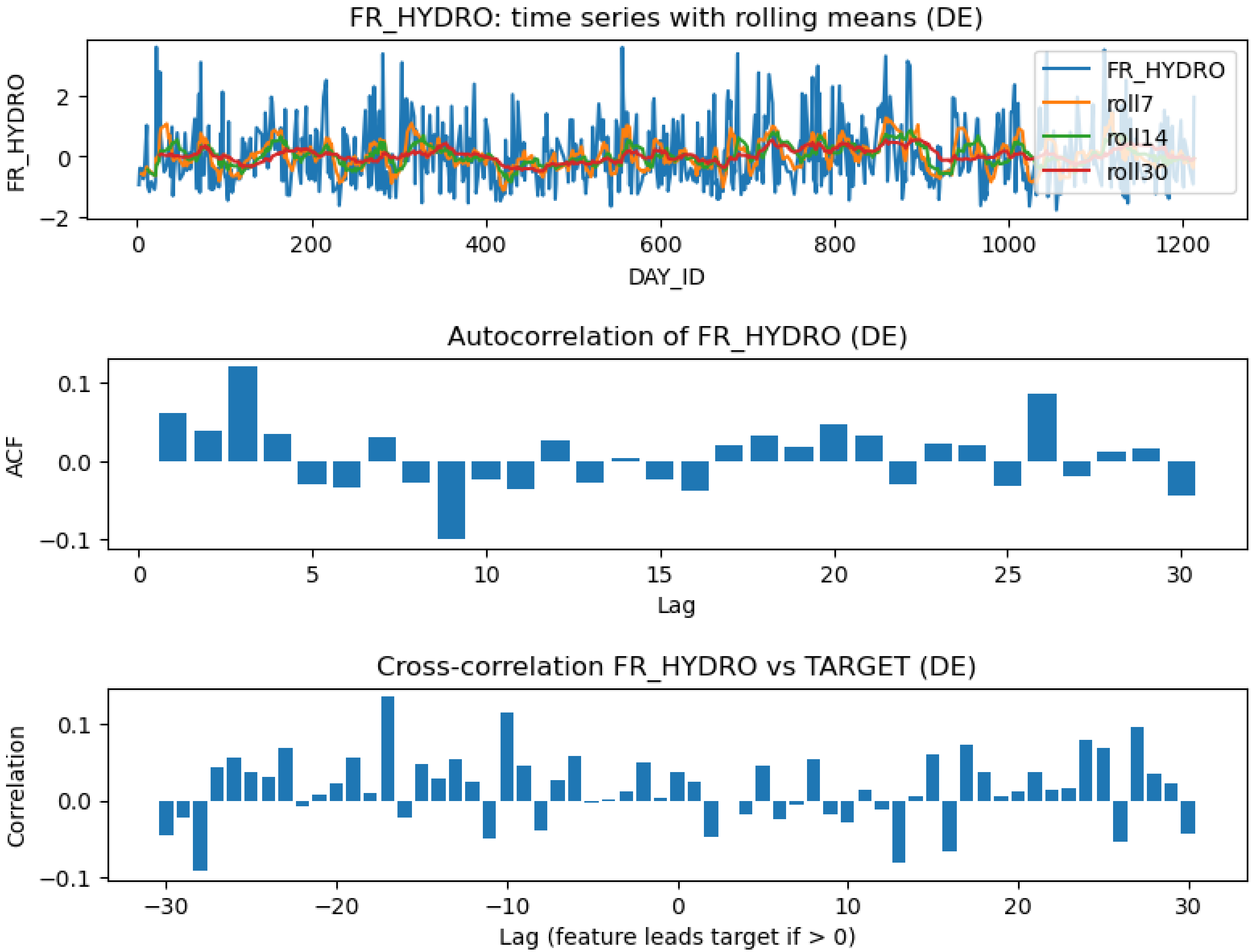

FR_GAS: time series with rolling means (DE)
FR_GAS
DAY_ID
FR_GAS
roll7
roll14
roll30
Autocorrelation of FR_GAS (DE)
ACF
Lag
Cross-correlation FR_GAS vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)


DE_TEMP: time series with rolling means (DE)
DE_TEMP
2.5
0.0
−2.5
0
200
400
600
800
1000
1200
DAY_ID
DE_TEMP
roll7
roll14
roll30
Autocorrelation of DE_TEMP (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag
Cross-correlation DE_TEMP vs TARGET (DE)
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

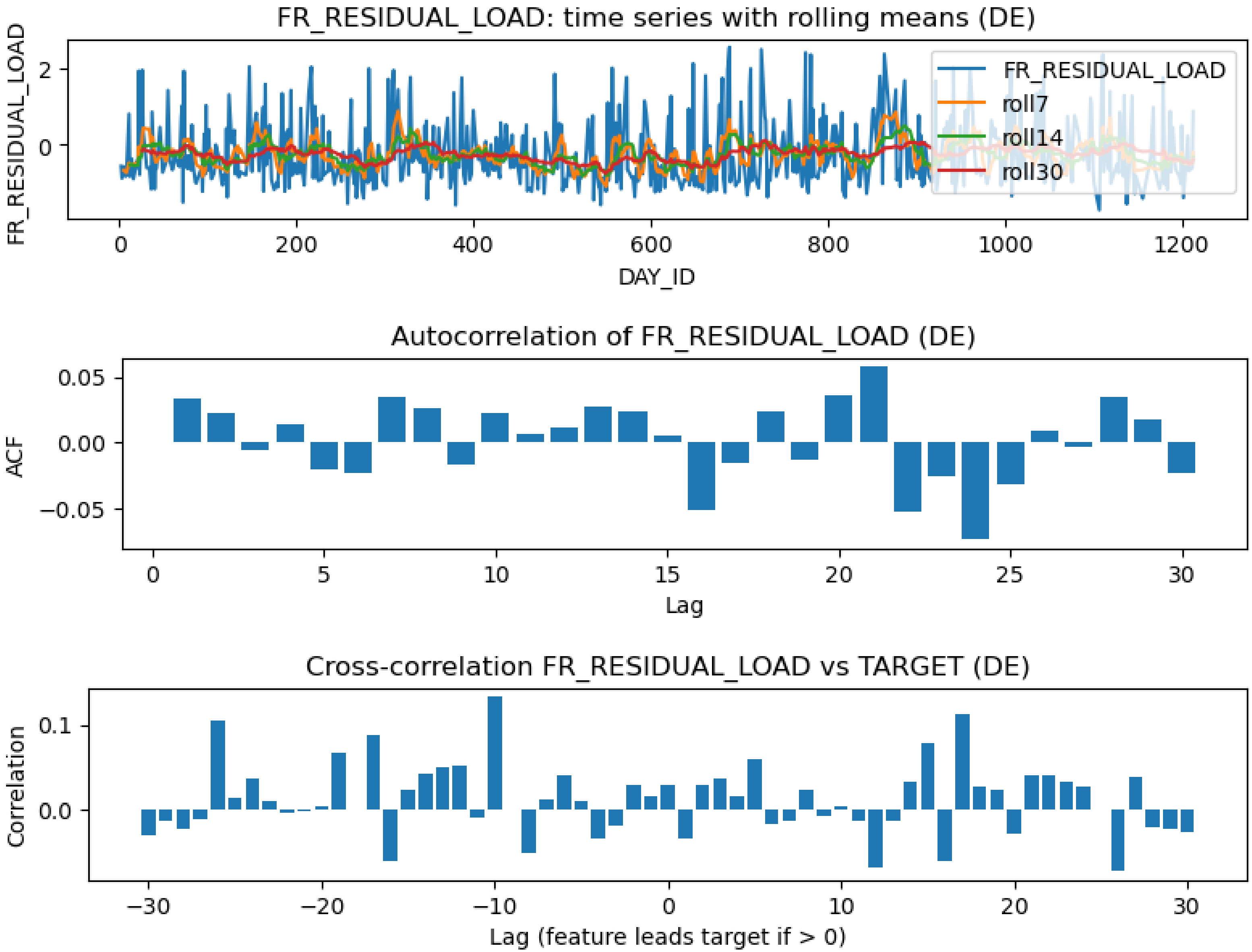

FR_RAIN: time series with rolling means (DE)
FR_RAIN
roll7
roll14
roll30
DAY_ID

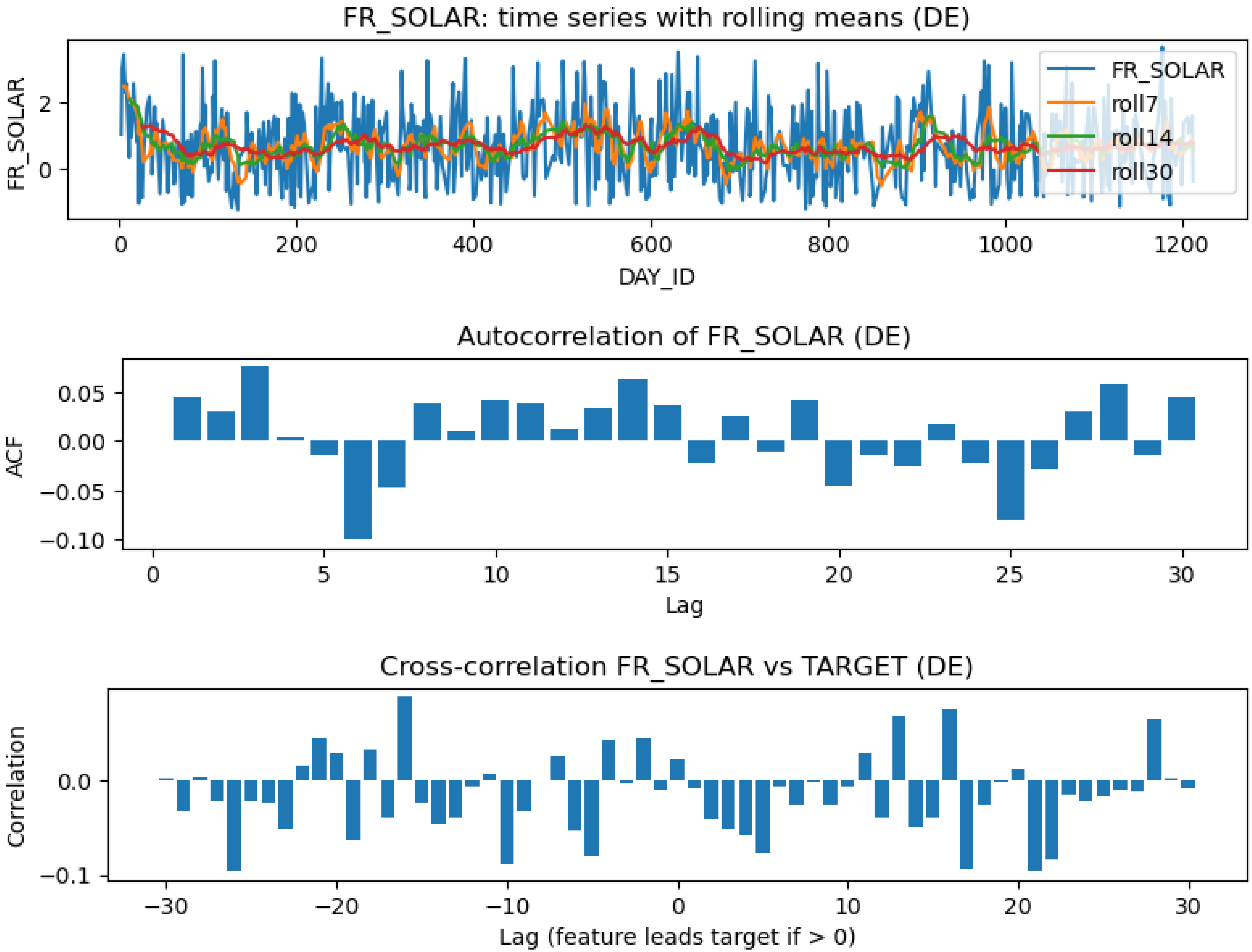

Autocorrelation of FR_RAIN (DE)
ACF
Lag

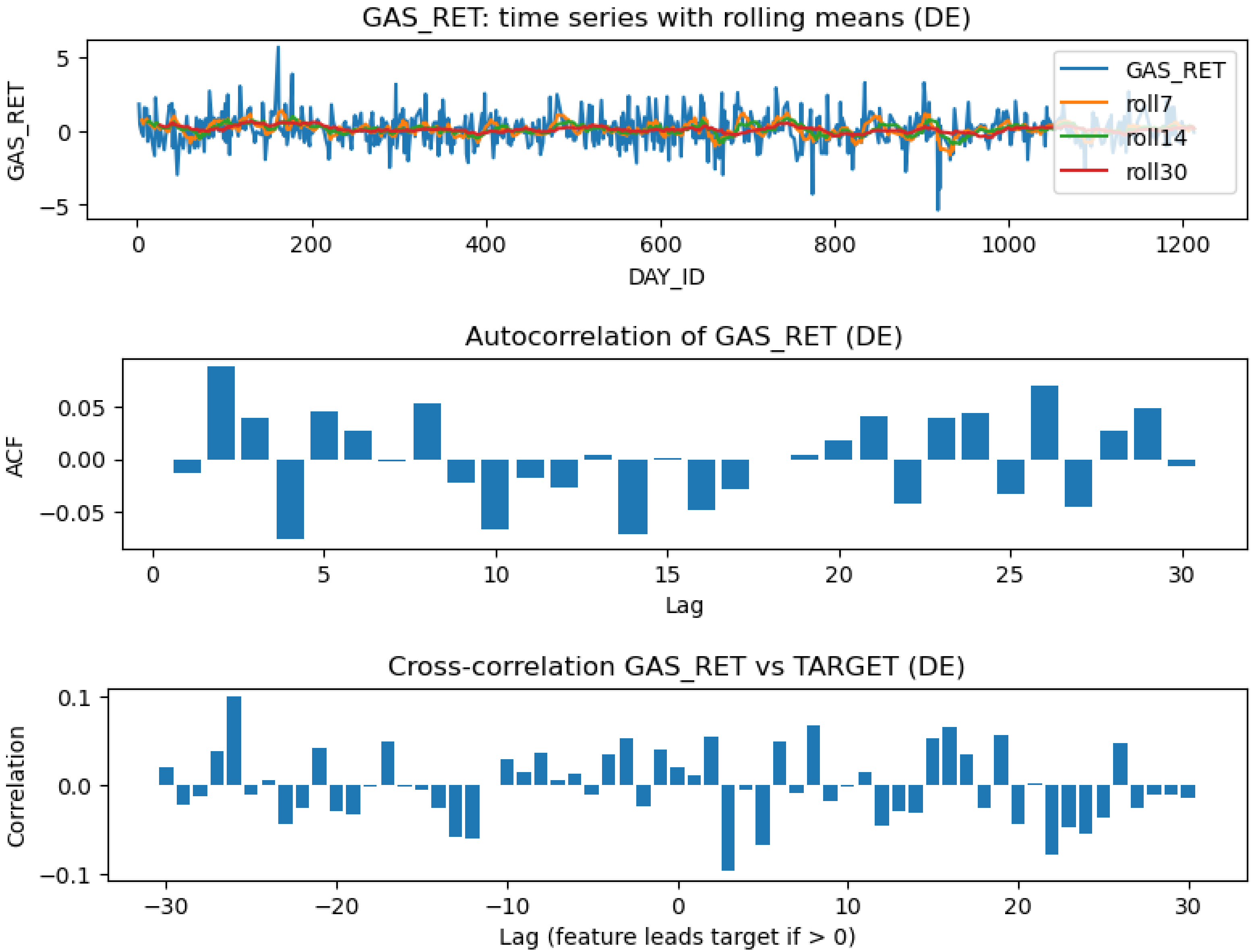

Cross-correlation FR_RAIN vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

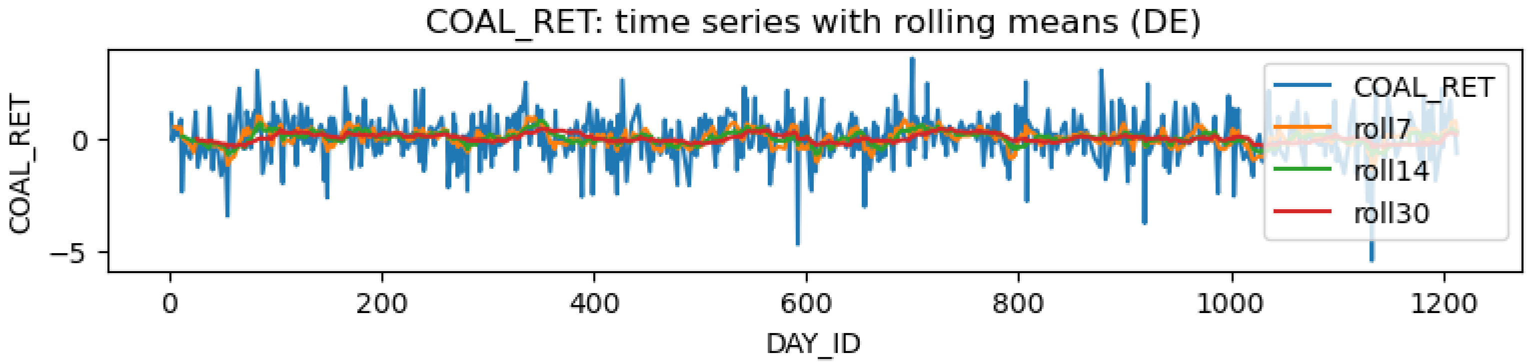


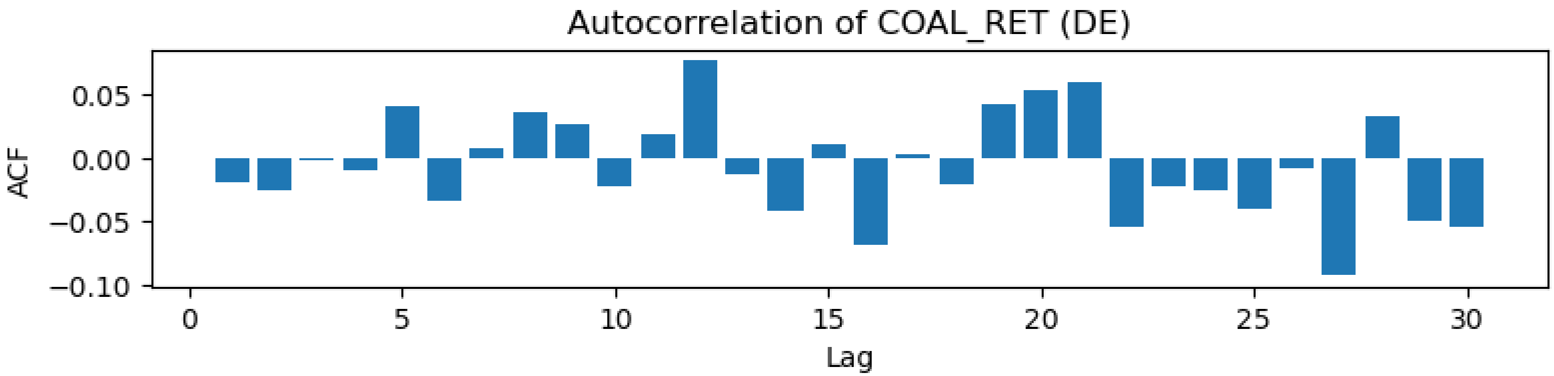
FR_HYDRO: time series with rolling means (DE)
FR_HYDRO
roll7
roll14
roll30
DAY_ID
Autocorrelation of FR_HYDRO (DE)
ACF
Lag
Cross-correlation FR_HYDRO vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

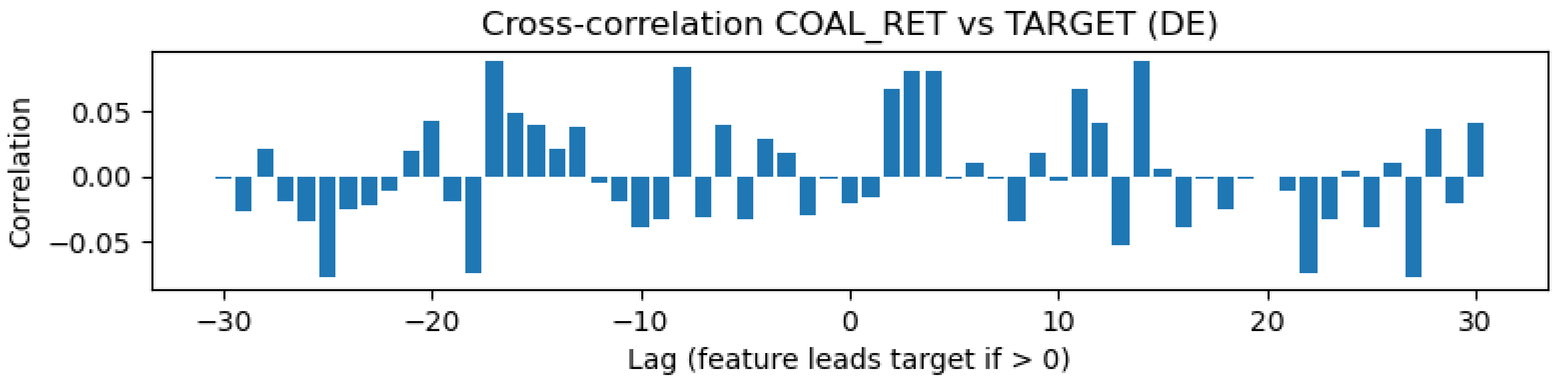

FR_RESIDUAL_LOAD: time series with rolling means (DE)
FR_RESIDUAL_LOAD
DAY_ID
FR_RESIDUAL_LOAD
roll7
roll14
roll30
Autocorrelation of FR_RESIDUAL_LOAD (DE)
ACF
Lag
Cross-correlation FR_RESIDUAL_LOAD vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

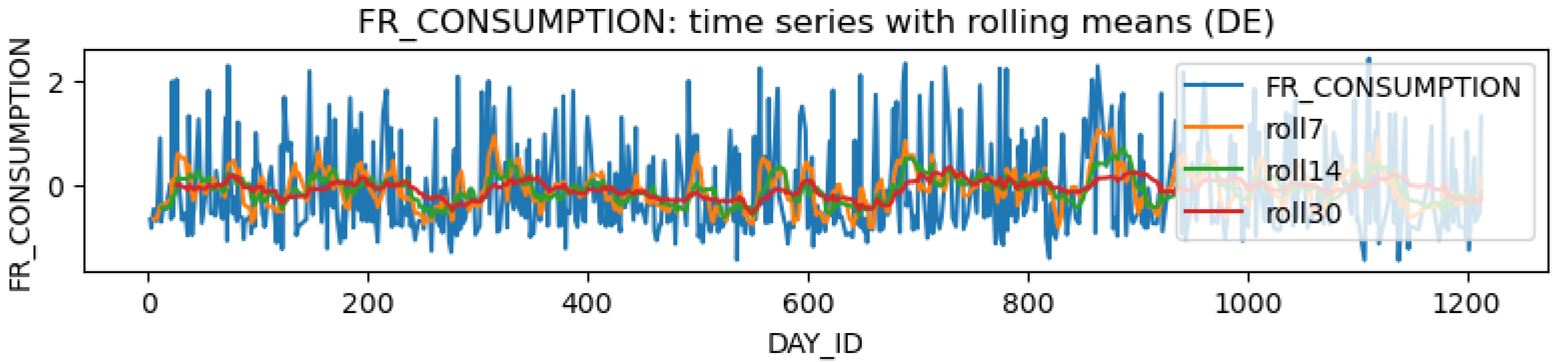

FR_SOLAR: time series with rolling means (DE)
FR_SOLAR
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_SOLAR
roll7
roll14
roll30
Autocorrelation of FR_SOLAR (DE)
ACF
0.05
0.00
−0.05
−0.10
0
5
10
15
20
25
30
Lag
Cross-correlation FR_SOLAR vs TARGET (DE)
Correlation
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

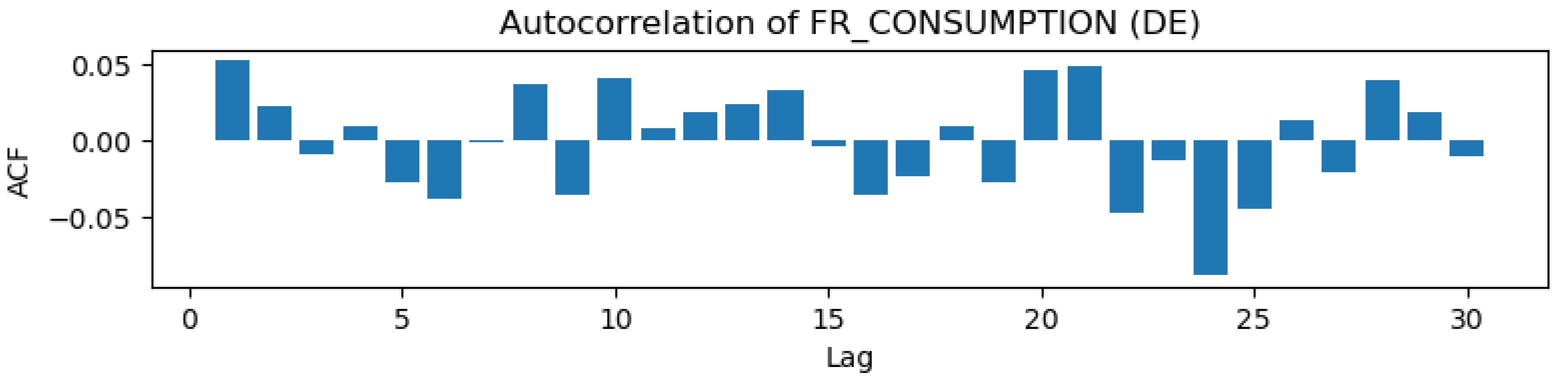

GAS_RET: time series with rolling means (DE)
GAS_RET
DAY_ID
GAS_RET
roll7
roll14
roll30
5
0
−5
0
200
400
600
800
1000
1200
Autocorrelation of GAS_RET (DE)
ACF
Lag
0.05
0.00
−0.05
0
5
10
15
20
25
30
Cross-correlation GAS_RET vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)
0.1
0.0
−0.1
−30
−20
−10
0
10
20
30

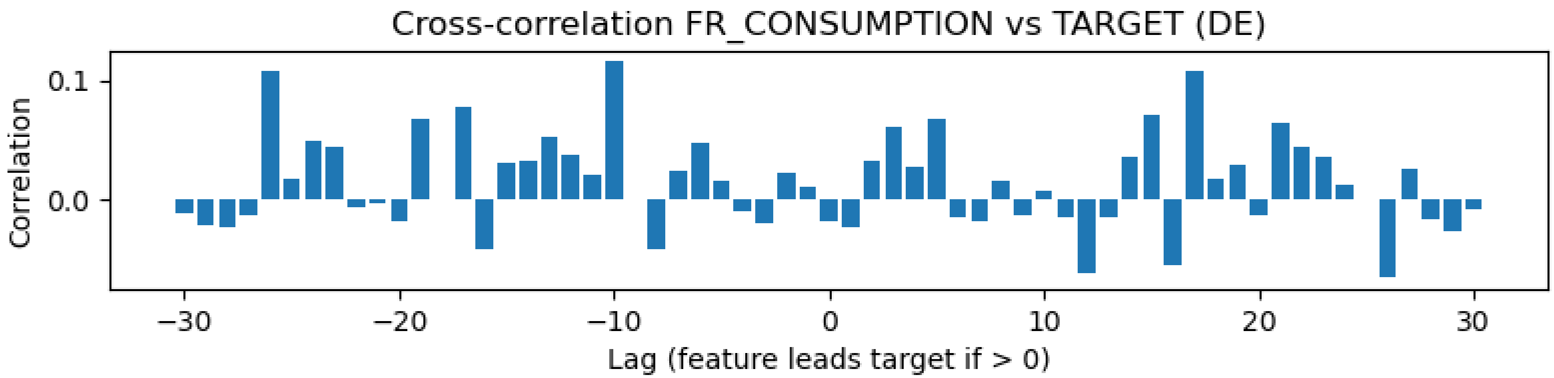

COAL_RET: time series with rolling means (DE)
COAL_RET
0
−5
0
200
400
600
800
1000
1200
DAY_ID
COAL_RET
roll7
roll14
roll30

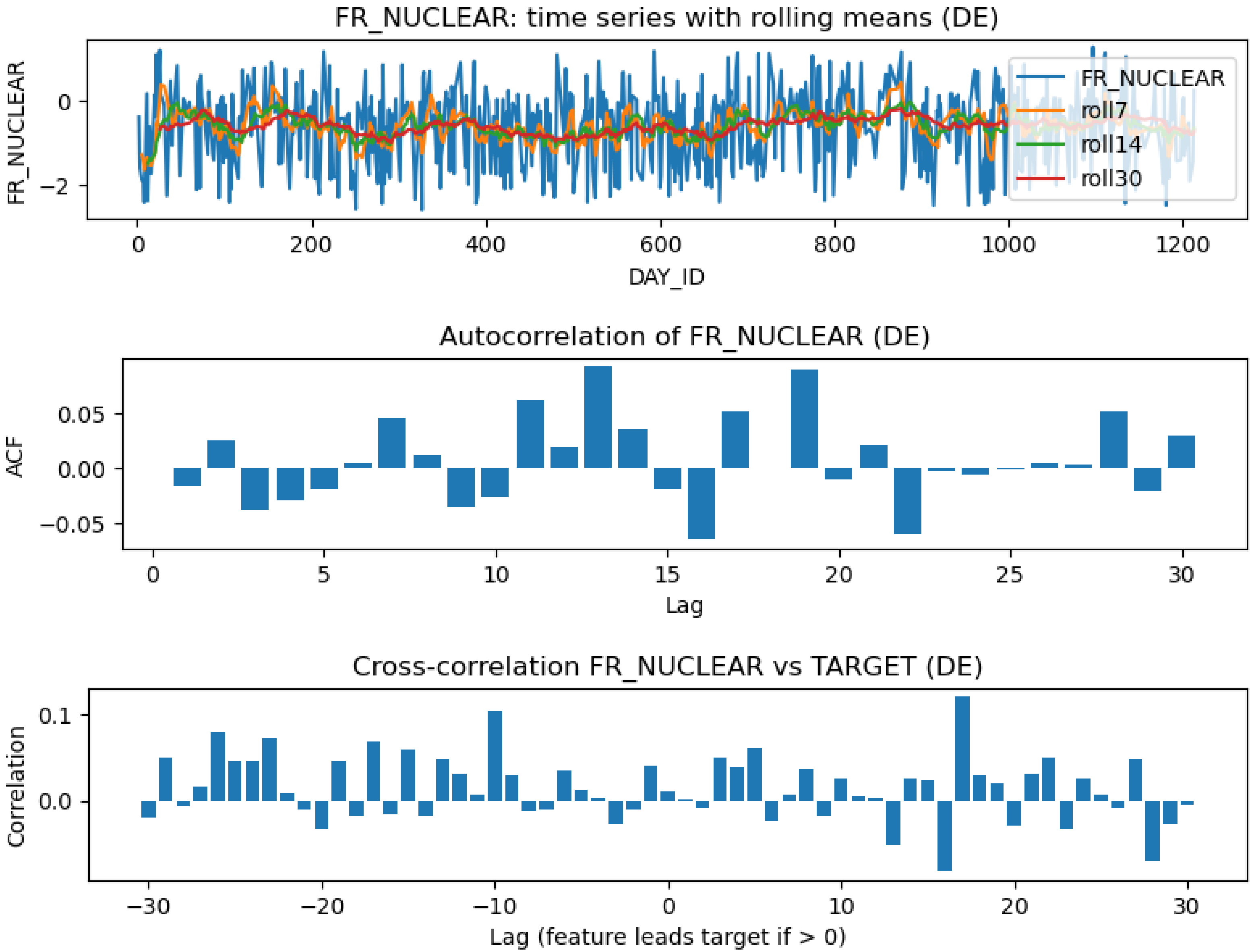

Autocorrelation of COAL_RET (DE)
ACF
0.05
0.00
−0.05
−0.10
0
5
10
15
20
25
30
Lag

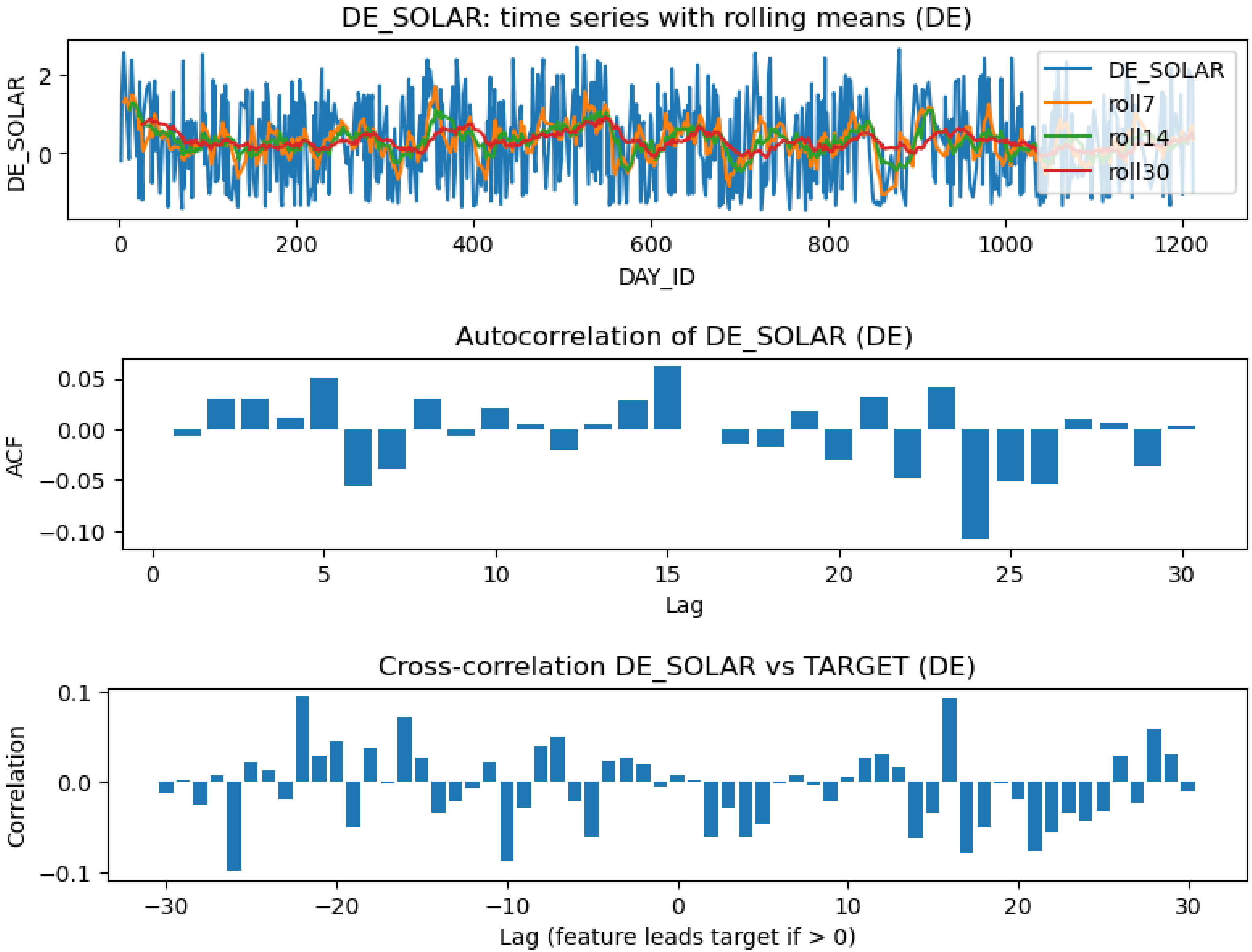

Cross-correlation COAL_RET vs TARGET (DE)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

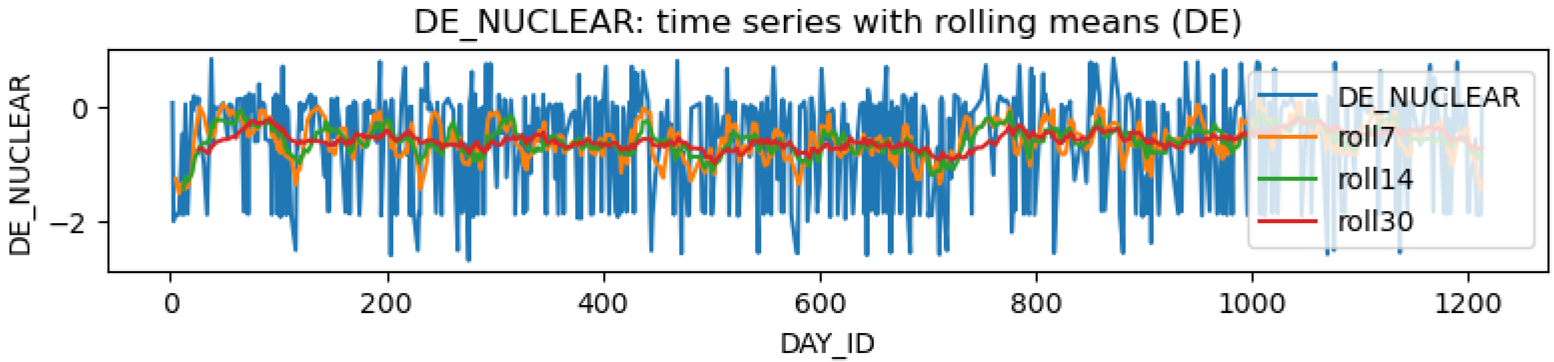

FR_CONSUMPTION: time series with rolling means (DE)
FR_CONSUMPTION
2
0
0
200
400
600
800
1000
1200
DAY_ID
FR_CONSUMPTION
roll7
roll14
roll30

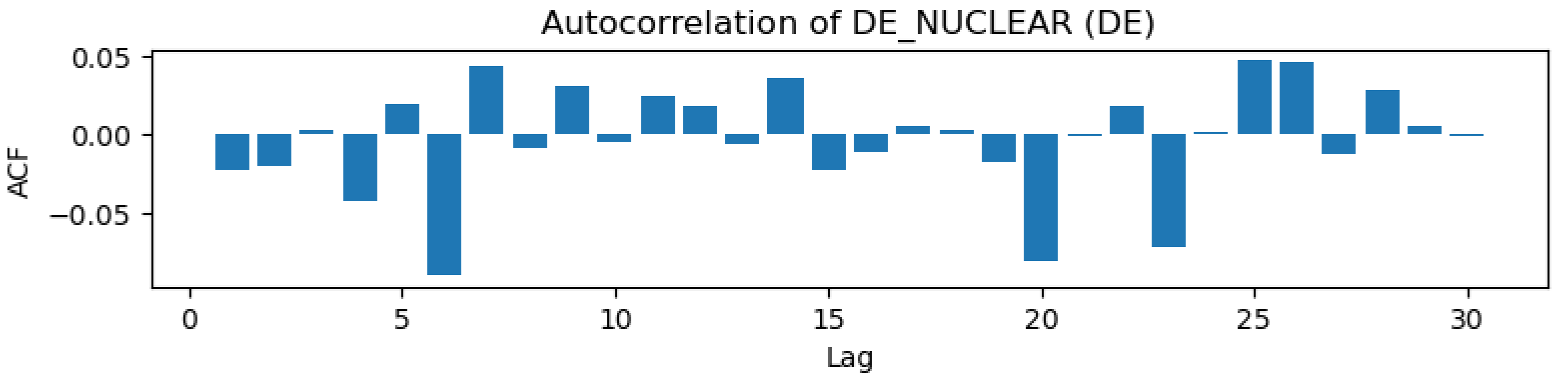

Autocorrelation of FR_CONSUMPTION (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

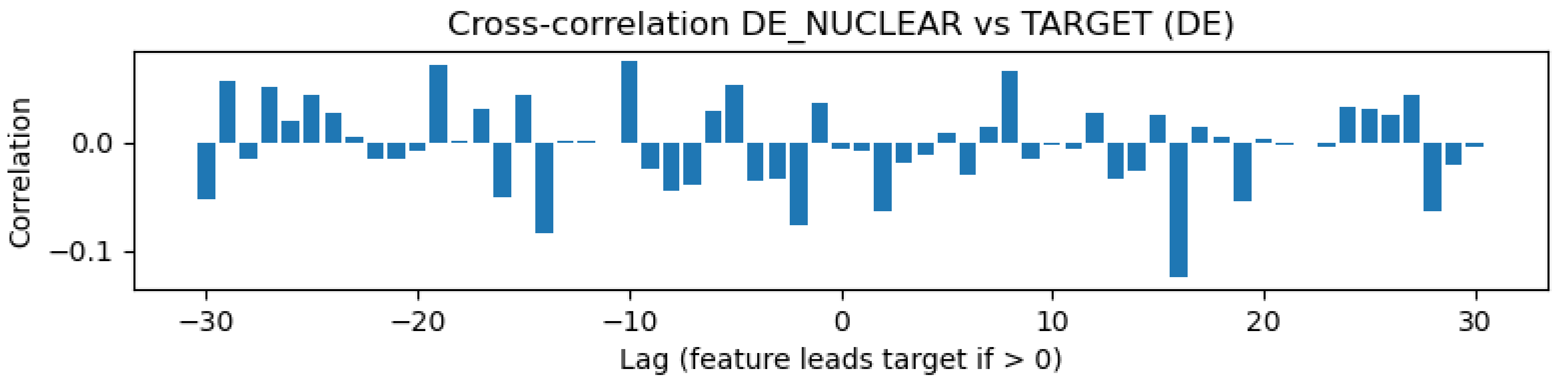

Cross-correlation FR_CONSUMPTION vs TARGET (DE)
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

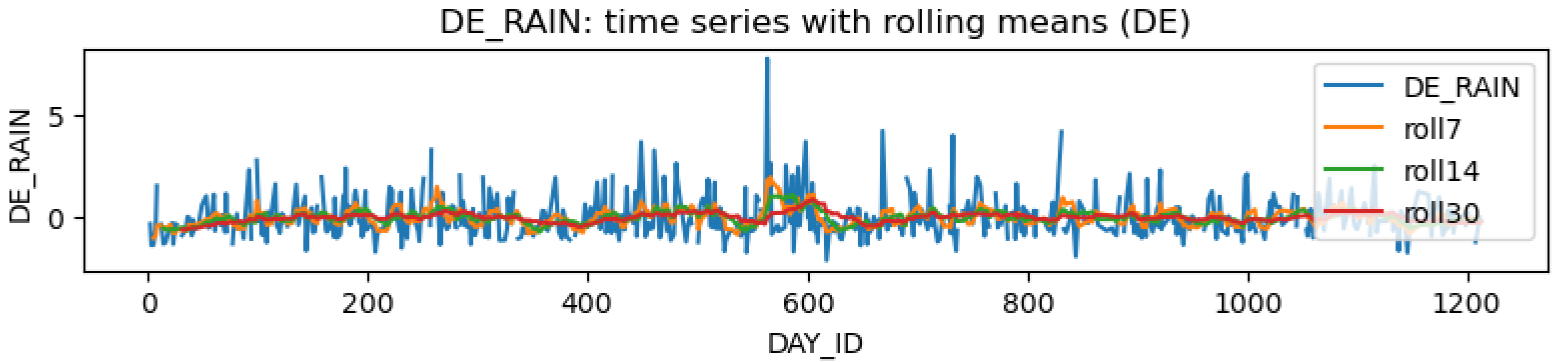

FR_NUCLEAR: time series with rolling means (DE)
FR_NUCLEAR
0
−2
0
200
400
600
800
1000
1200
DAY_ID
FR_NUCLEAR
roll7
roll14
roll30
Autocorrelation of FR_NUCLEAR (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag
Cross-correlation FR_NUCLEAR vs TARGET (DE)
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

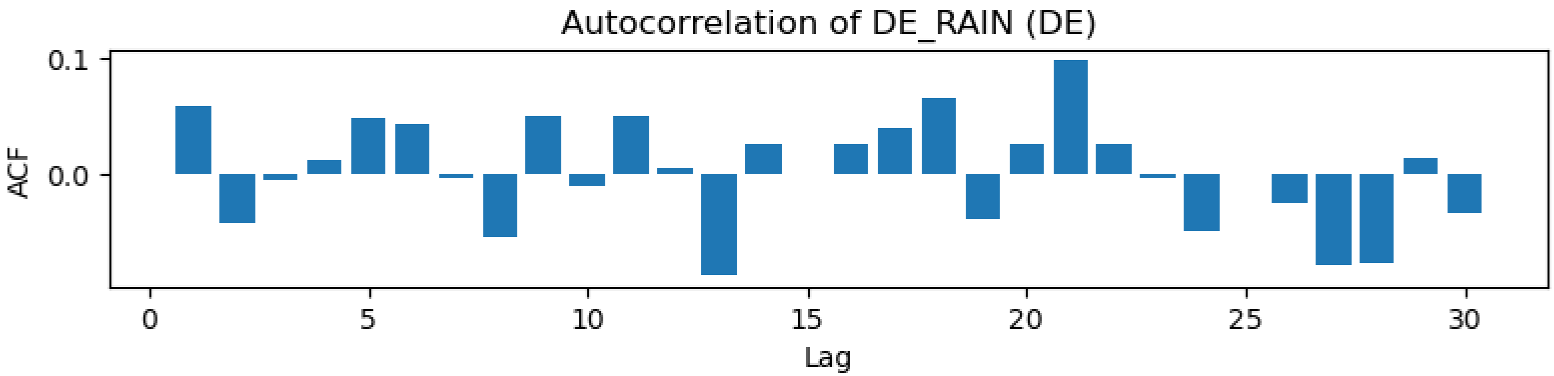

DE_SOLAR: time series with rolling means (DE)
DE_SOLAR
roll7
roll14
roll30
DE_SOLAR
DAY_ID
Autocorrelation of DE_SOLAR (DE)
ACF
Lag
Cross-correlation DE_SOLAR vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

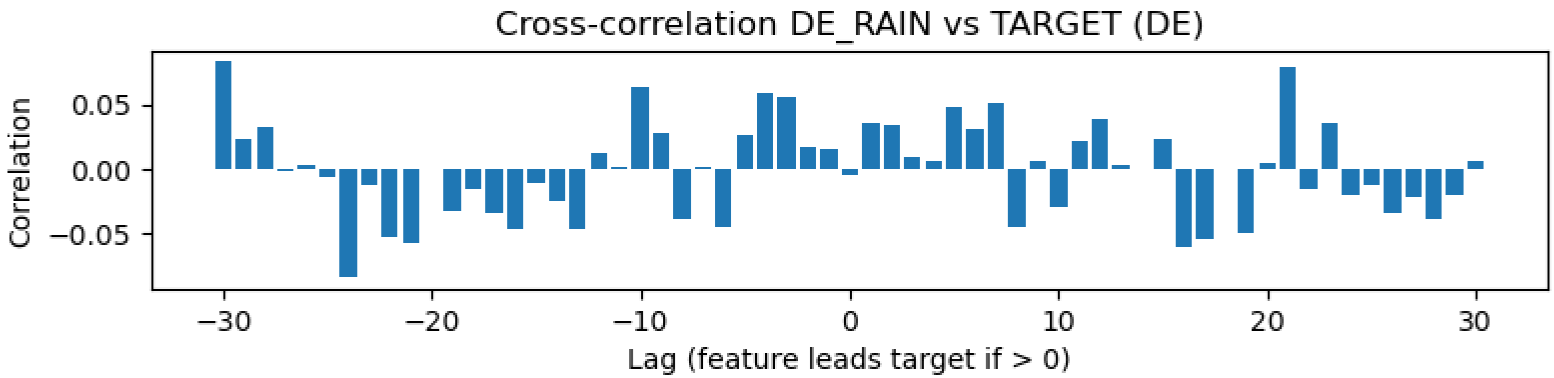

DE_NUCLEAR: time series with rolling means (DE)
DE_NUCLEAR
0
−2
0
200
400
600
800
1000
1200
DAY_ID
DE_NUCLEAR
roll7
roll14
roll30

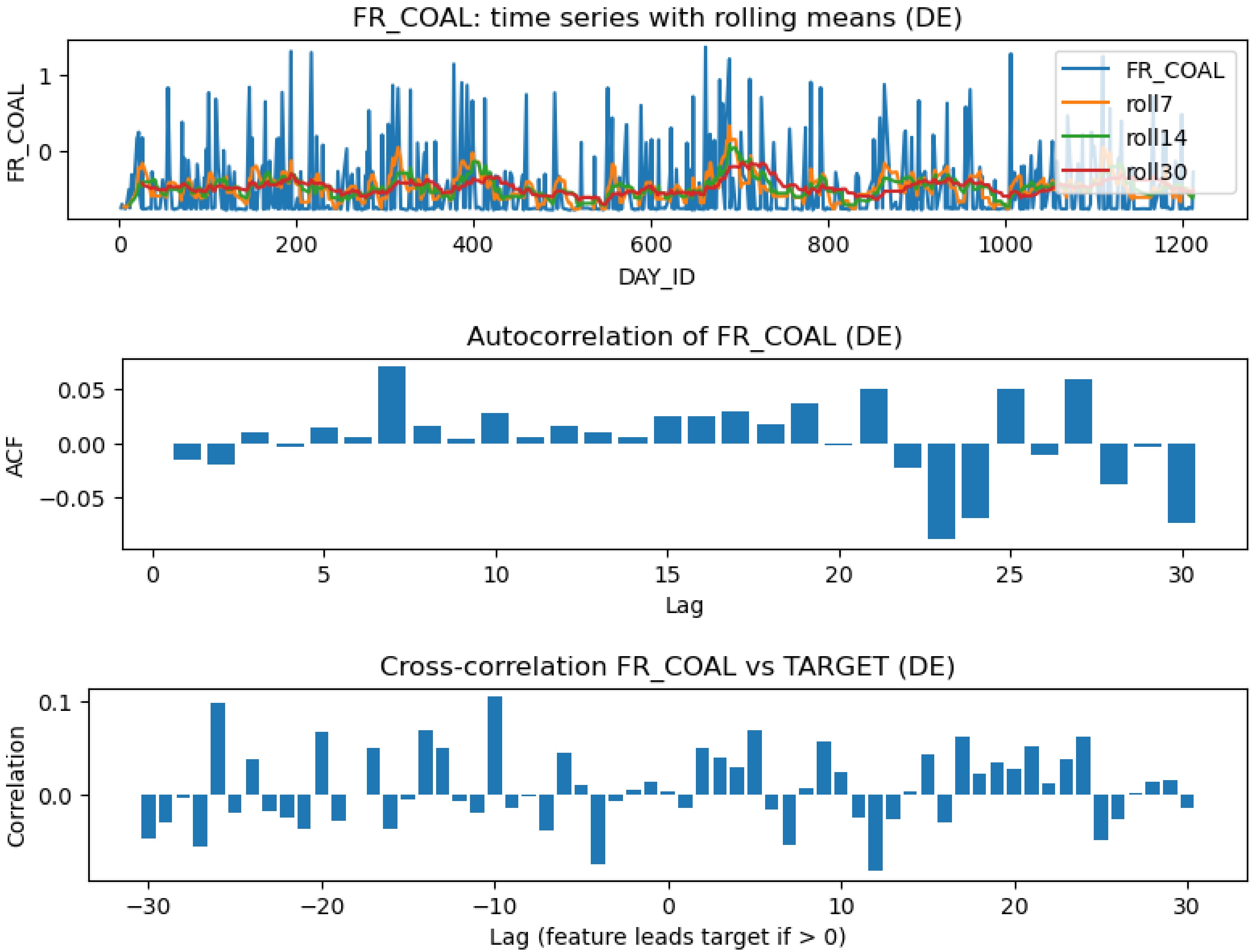

Autocorrelation of DE_NUCLEAR (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

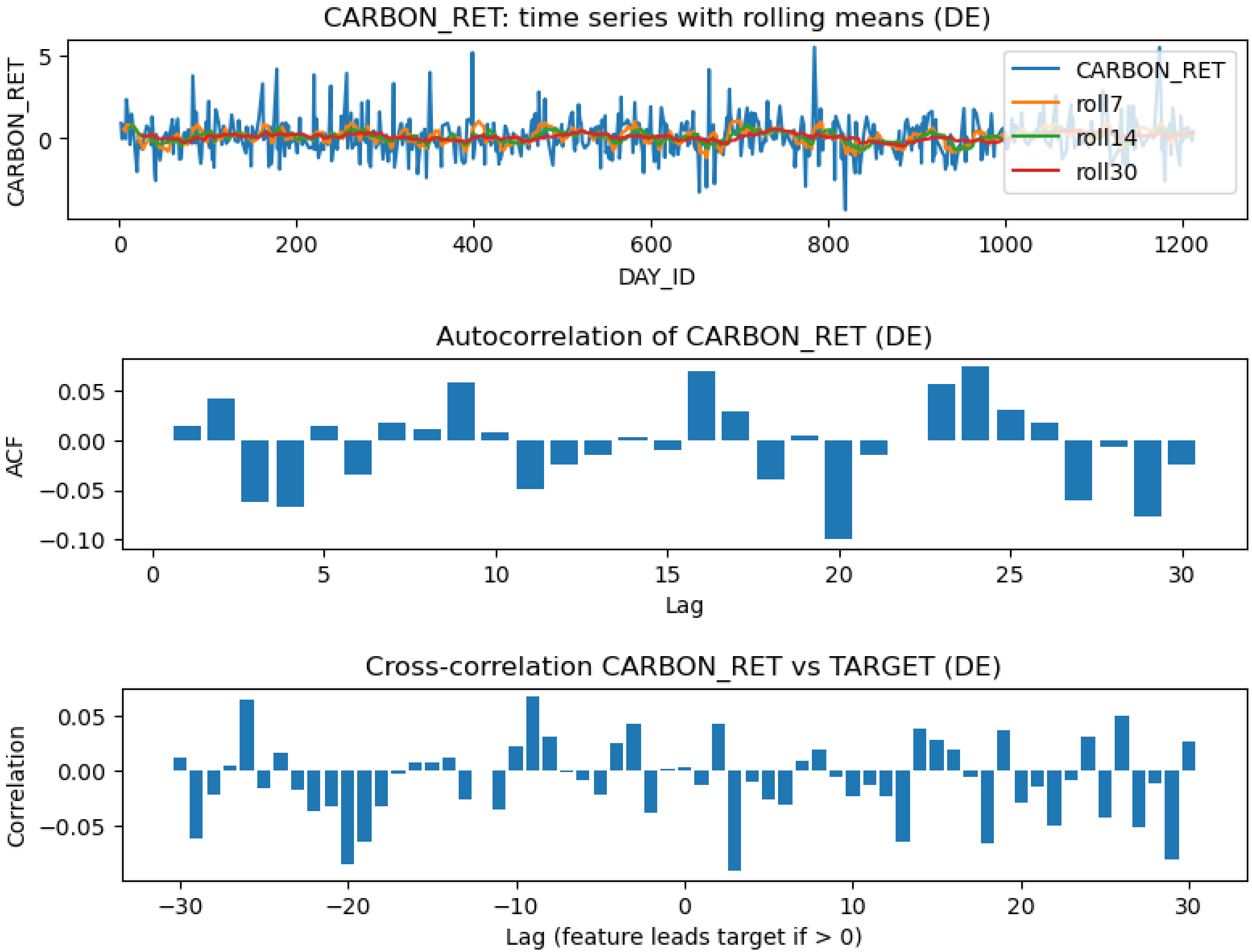

Cross-correlation DE_NUCLEAR vs TARGET (DE)
Correlation
0.0
−0.1
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)


DE_RAIN: time series with rolling means (DE)
DE_RAIN
roll7
roll14
roll30
DE_RAIN
0
5
0
200
400
600
800
1000
1200
DAY_ID

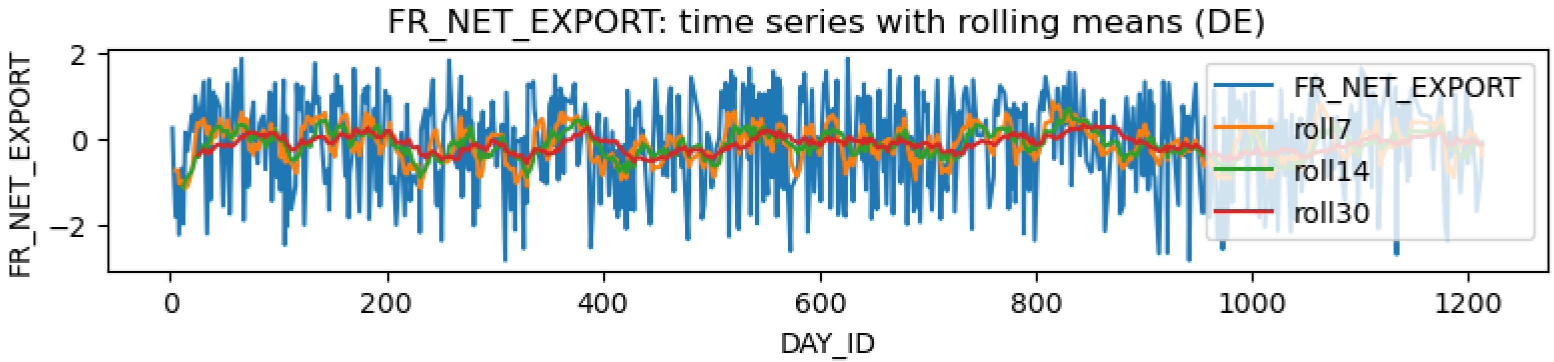

Autocorrelation of DE_RAIN (DE)
ACF
0.1
0.0
0
5
10
15
20
25
30
Lag

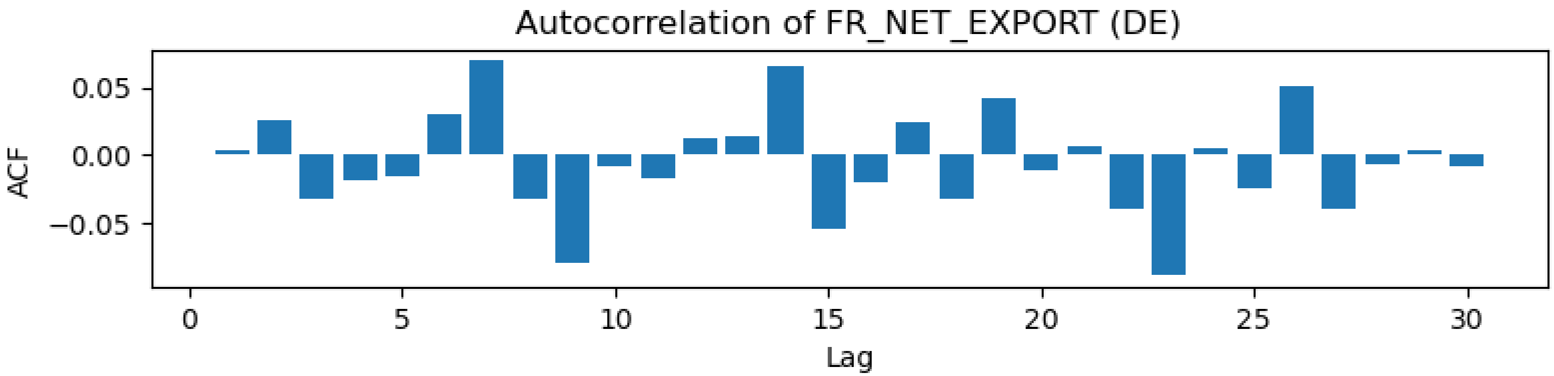

Cross-correlation DE_RAIN vs TARGET (DE)
Correlation
0.05
0.00
−0.05
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)

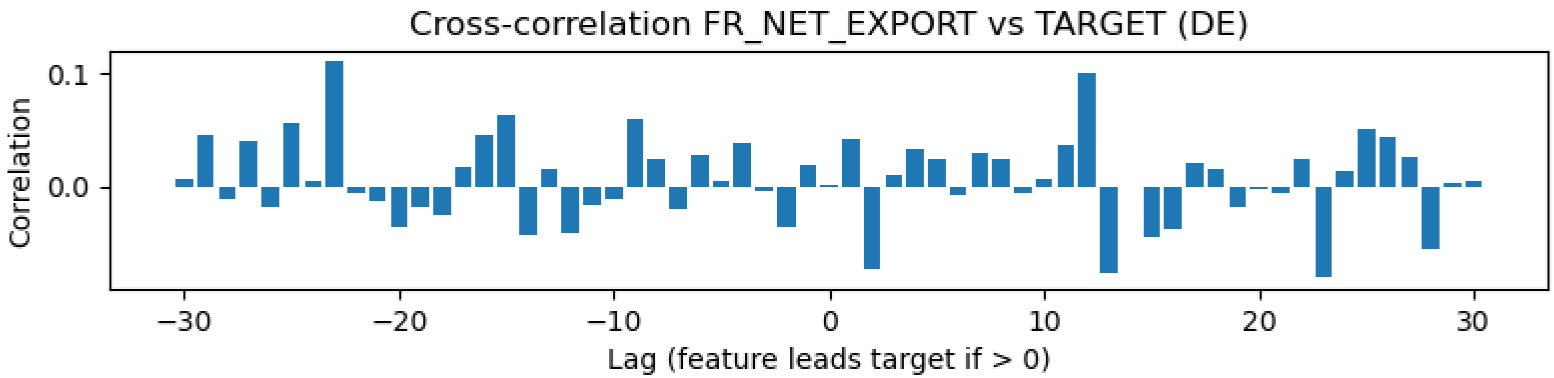

FR_COAL: time series with rolling means (DE)
FR_COAL
roll7
roll14
roll30
FR_COAL
DAY_ID
0
200
400
600
800
1000
1200
Autocorrelation of FR_COAL (DE)
ACF
0.05
0.00
−0.05
Lag
Cross-correlation FR_COAL vs TARGET (DE)
Correlation
0.1
0.0
Lag (feature leads target if > 0)

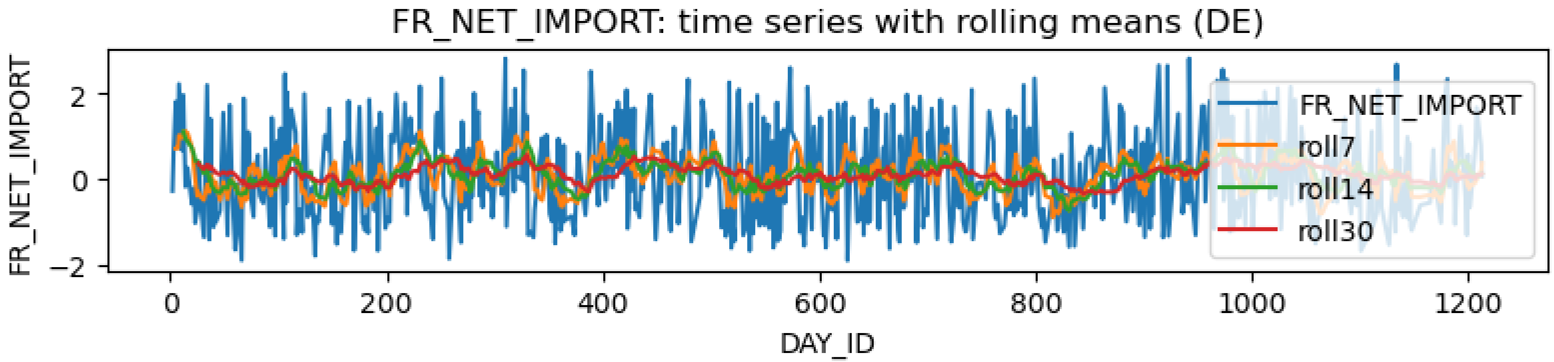


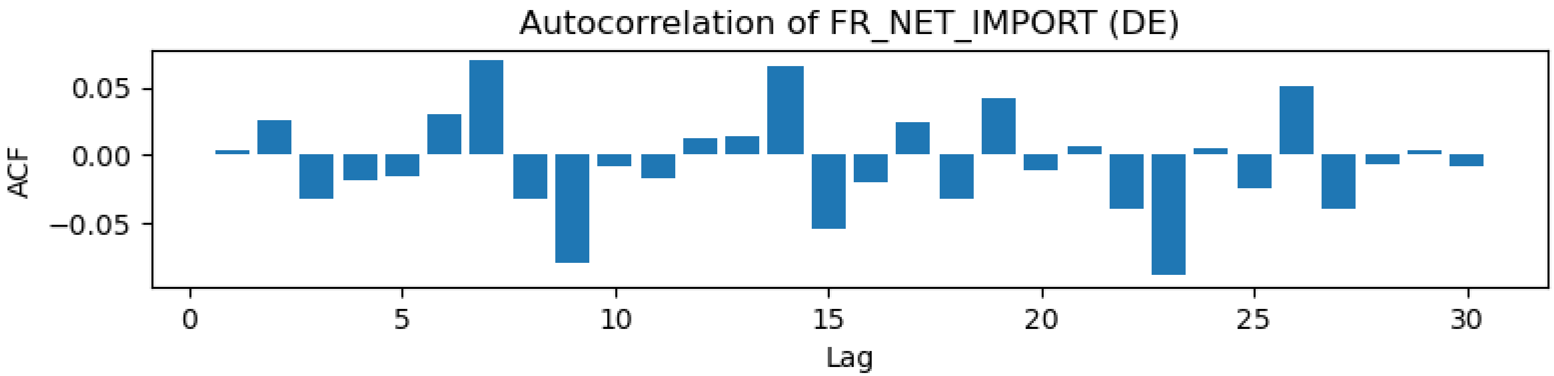
CARBON_RET: time series with rolling means (DE)
CARBON_RET
roll7
roll14
roll30
DAY_ID
Autocorrelation of CARBON_RET (DE)
ACF
Lag
Cross-correlation CARBON_RET vs TARGET (DE)
Correlation
Lag (feature leads target if > 0)

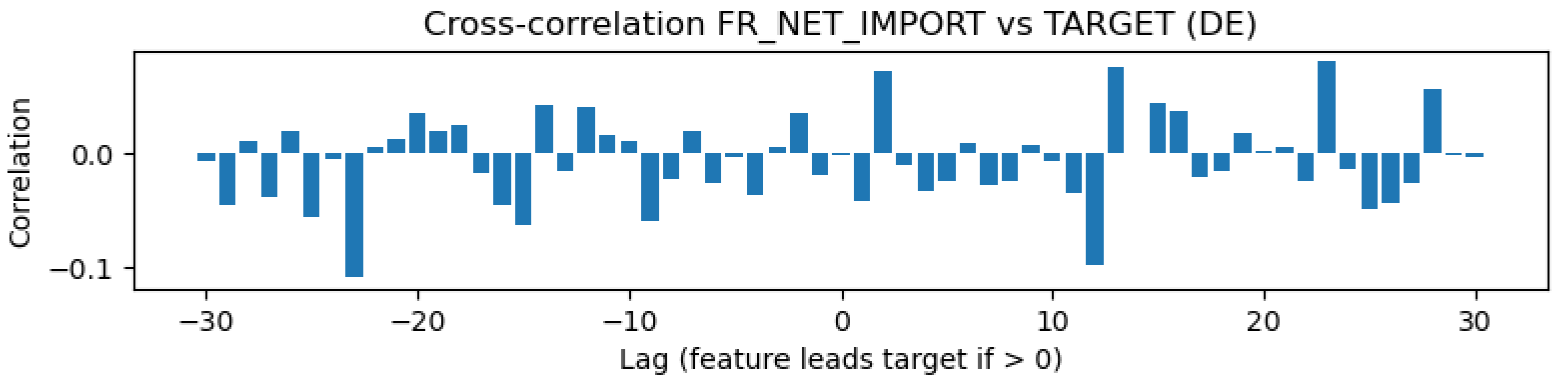

FR_NET_EXPORT: time series with rolling means (DE)
FR_NET_EXPORT
2
0
−2
0
200
400
600
800
1000
1200
DAY_ID
FR_NET_EXPORT
roll7
roll14
roll30


Autocorrelation of FR_NET_EXPORT (DE)
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag


Cross-correlation FR_NET_EXPORT vs TARGET (DE)
Correlation
0.1
0.0
−30
−20
−10
0
10
20
30
Lag (feature leads target if > 0)







10. Missing Value Imputation Comparison

In [19]:
```
features = num_cols.copy()
y_vec = df["TARGET"].to_numpy()

def quick_eval(X_mat, y_vec):
    if "DAY_ID" in df.columns:
        order = np.argsort(df["DAY_ID"].to_numpy())
        split = int(0.8 * len(order))
        tr_idx, va_idx = order[:split], order[split:]
    else:
        tr_idx, va_idx = train_test_split(np.arange(len(df)), test_size=0.2, random_state=RANDOM_STATE)

    model = HistGradientBoostingRegressor(random_state=RANDOM_STATE, max_iter=600)
    model.fit(X_mat[tr_idx], y_vec[tr_idx])
    pred = model.predict(X_mat[va_idx])
    rmse = np.sqrt(np.mean((pred - y_vec[va_idx])**2))
    return rmse

imp_med = SimpleImputer(strategy="median")
X_med = imp_med.fit_transform(df[features])
rmse_med = quick_eval(X_med, y_vec)

imp_knn = KNNImputer(n_neighbors=5, weights="distance")
X_knn = imp_knn.fit_transform(df[features])
rmse_knn = quick_eval(X_knn, y_vec)

print({"rmse_median": rmse_med, "rmse_knn": rmse_knn})
use_knn = rmse_knn <= rmse_med
X_imputed = X_knn if use_knn else X_med
print("Chosen imputer:", "KNNImputer" if use_knn else "SimpleImputer(median)")
```

```
{'rmse_median': np.float64(1.034780231542586), 'rmse_knn': np.float64(1.0444992507742312)}
Chosen imputer: SimpleImputer(median)
```

11. Feature Selection

In [20]:
```
print("Before Feature Selection:", len(features))
vt = VarianceThreshold(threshold=1e-8)
X_vt = vt.fit_transform(X_imputed)
feats_vt = [f for f, keep in zip(features, vt.get_support()) if keep]
```

```
print("Kept after VarianceThreshold:", len(feats_vt))

def correlation_filter(X_df, cols, threshold=0.99):
    C = X_df[cols].corr().abs()
    keep = []
    dropped = set()
    for c in cols:
        if c in dropped:
            continue
        keep.append(c)
        high = C.index[(C[c] > threshold) & (C.index != c)].tolist()
        for h in high:
            dropped.add(h)
    return keep, sorted(list(dropped))

keep_cols, dropped_cols = correlation_filter(pd.DataFrame(X_imputed, columns=features), feats_vt, threshold=0.99)
print("Dropped due to high correlation:", dropped_cols)
print("Remaining after correlation filter:", len(keep_cols))
```

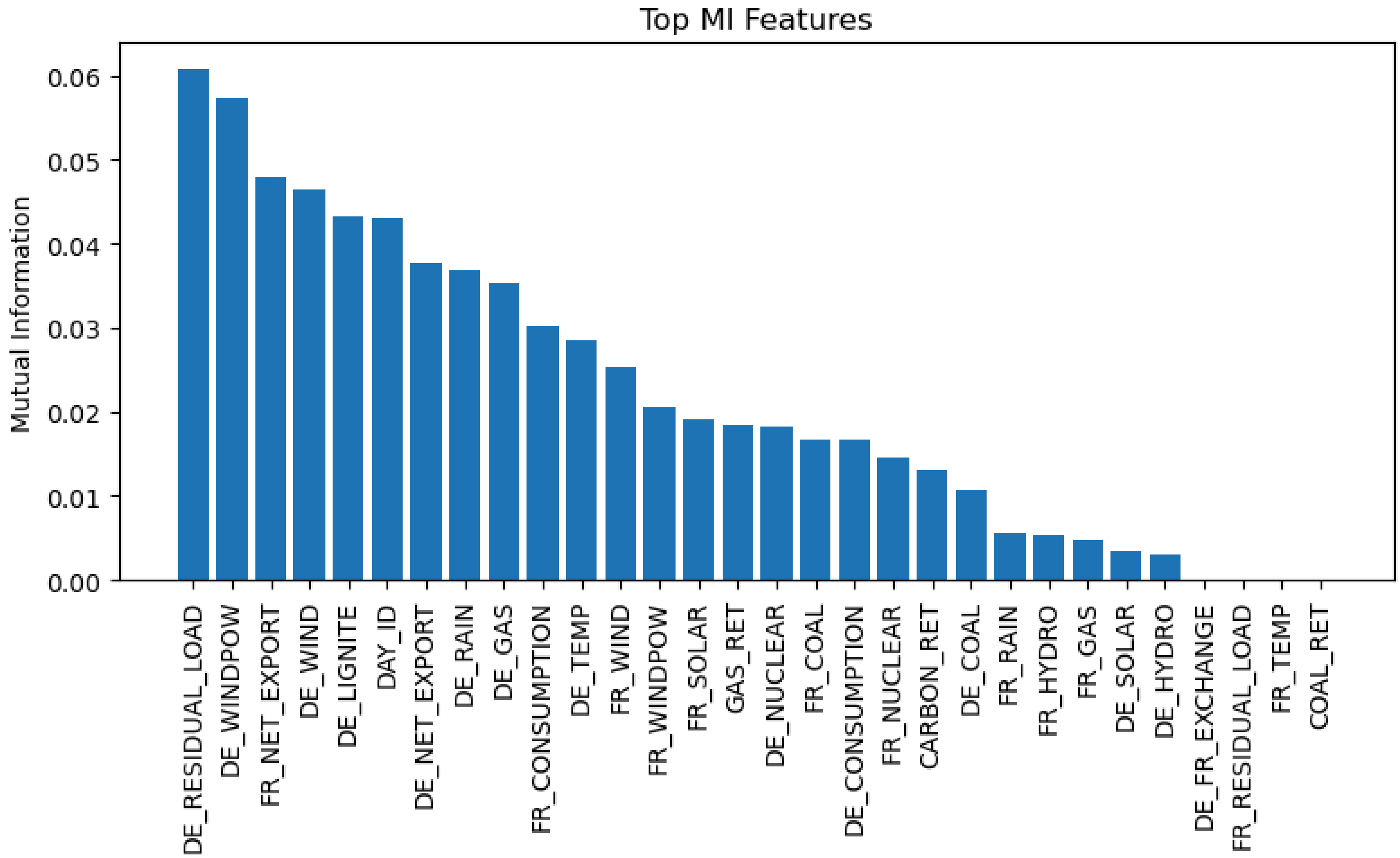
```
Before Feature Selection: 33
Kept after VarianceThreshold: 33
Dropped due to high correlation: ['DE_NET_IMPORT', 'FR_DE_EXCHANGE', 'FR_NET_IMPORT']
Remaining after correlation filter: 30
```

12. Mutual Information

In [21]:

```
# Mutual Information ranking (top-k)
mi = mutual_info_regression(pd.DataFrame(X_imputed, columns=features)[keep_cols], y_vec, random_state=RANDOM_STATE)
mi_series = pd.Series(mi, index=keep_cols).sort_values(ascending=False)
display(mi_series)

plt.figure()
plt.bar(mi_series.index.astype(str), mi_series.values)
plt.xticks(rotation=90)
plt.ylabel("Mutual Information")
plt.title("Top MI Features")
plt.tight_layout()
plt.show()
```

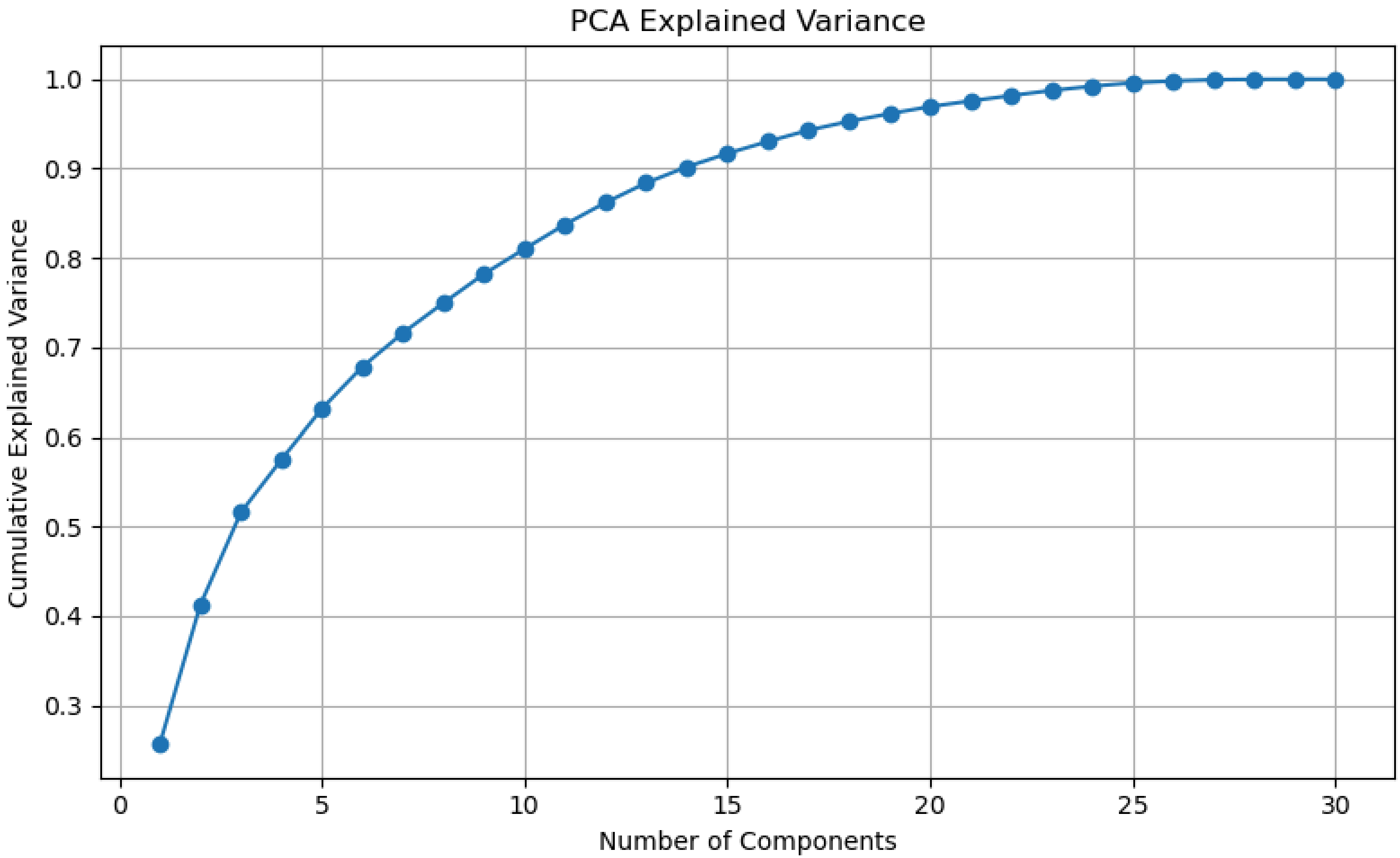
```
DE_RESIDUAL_LOAD    0.060875
DE_WINDPOW          0.057345
FR_NET_EXPORT       0.047978
DE_WIND             0.046520
DE_LIGNITE          0.043323
DAY_ID              0.042981
DE_NET_EXPORT       0.037728
DE_RAIN             0.036929
DE_GAS              0.035325
FR_CONSUMPTION      0.030100
DE_TEMP             0.028471
FR_WIND             0.025232
FR_WINDPOW          0.020598
FR_SOLAR            0.019060
GAS_RET             0.018451
DE_NUCLEAR          0.018269
FR_COAL             0.016737
DE_CONSUMPTION      0.016692
FR_NUCLEAR          0.014526
CARBON_RET          0.013019
DE_COAL             0.010716
FR_RAIN             0.005695
FR_HYDRO            0.005402
FR_GAS              0.004800
DE_SOLAR            0.003395
DE_HYDRO            0.003095
DE_FR_EXCHANGE      0.000000
FR_RESIDUAL_LOAD    0.000000
FR_TEMP             0.000000
COAL_RET            0.000000
dtype: float64
```



13. Feature Extraction (PCA)

In [22]:
```
scaler = StandardScaler()
X_scaled = scaler.fit_transform(pd.DataFrame(X_imputed, columns=features)[keep_cols])

pca = PCA(n_components=X_scaled.shape[1], random_state=RANDOM_STATE)
X_pca = pca.fit_transform(X_scaled)

cum_var = np.cumsum(pca.explained_variance_ratio_)
```

```python
plt.figure()
plt.plot(np.arange(1, len(cum_var)+1), cum_var, marker='o')
plt.xlabel("Number of Components")
plt.ylabel("Cumulative Explained Variance")
plt.title("PCA Explained Variance")
plt.grid(True)
plt.tight_layout()
plt.show()

print("Components to reach 95% variance:", int((cum_var >= 0.95).argmax()+1) if (cum_var >= 0.95).any() else len(cum_var))
```




```
Components to reach 95% variance: 18
```

In [23]:
```
if "COUNTRY" in df.columns and df["COUNTRY"].nunique()<=10:
    plt.figure()
    tmp=df.copy(); tmp["PC1"]=X_pca[:,0]; tmp["PC2"]=X_pca[:,1]
    for k,g in tmp.groupby("COUNTRY"):
        plt.scatter(g["PC1"],g["PC2"],alpha=0.6,label=str(k))
    plt.legend(); plt.title("PCA (PC1 vs PC2) by COUNTRY"); plt.tight_layout(); plt.show()
```