In [1]:
```
import pandas as pd
from darts import TimeSeries
```

In [2]:
```
df = pd.read_csv("MY3_May_2023_KNN_Imputed.csv")

df
```

Out[2]:

	Datetime	id	field1	field2	field3	field4	field5	field6	field7	field8
0	2022-07-03 20:30:00	2.200000e+01	8.871990	0.000000	398.496241	3.636364	26.00000	71.000000	45.666667	55.000000
1	2022-07-03 20:45:00	6.700000e+01	3.802281	0.000000	229.323308	0.000000	25.00000	62.000000	44.666667	53.333333
2	2022-07-03 21:00:00	1.120000e+02	12.674271	0.000000	184.210526	0.000000	25.00000	60.000000	43.666667	52.750000
3	2022-07-03 21:15:00	1.560000e+02	16.476553	0.000000	165.413534	0.000000	24.00000	59.666667	43.333333	52.666667
4	2022-07-03 21:30:00	5.307278e+05	321.540266	190.816359	794.429588	371.997558	27.25921	61.186420	18.749892	20.904609
...	...	...	...	...	...	...	...	...	...	...
28042	2023-04-21 23:00:00	1.061851e+06	397.944200	225.108225	700.960219	489.898990	29.00000	69.000000	27.666667	29.000000
28043	2023-04-21 23:15:00	1.061896e+06	392.070485	225.974026	707.818930	492.424242	29.00000	69.000000	31.666667	36.000000
28044	2023-04-21 23:30:00	1.061941e+06	393.538913	225.974026	716.049383	493.686869	29.00000	70.000000	33.333333	39.333333
28045	2023-04-21 23:45:00	1.061986e+06	392.070485	225.108225	727.023320	488.636364	29.00000	70.000000	35.333333	42.666667
28046	2023-04-22 00:00:00	1.062008e+06	414.096916	220.779221	720.164609	496.212121	29.00000	70.000000	37.000000	44.000000

28047 rows × 10 columns

In [3]:
```
df1 = df
df1['Datetime'] = pd.to_datetime(df['Datetime'])

df1 = df.set_index('Datetime')
#df1
df2=df1.resample('1H').mean()
df2.reset_index(inplace = True)
df2
```

Out[3]:

	Datetime	id	field1	field2	field3	field4	field5	field6	field7	field8
0	2022-07-03 20:00:00	4.450000e+01	6.337136	0.000000	313.909774	1.818182	25.500000	66.500000	45.166667	54.166667
1	2022-07-03 21:00:00	2.654309e+05	168.057839	95.408180	484.620809	185.998779	25.879605	60.509877	31.124946	36.806471
2	2022-07-03 22:00:00	5.307278e+05	321.540266	190.816359	794.429588	371.997558	27.259210	61.186420	18.749892	20.904609
3	2022-07-03 23:00:00	5.307278e+05	321.540266	190.816359	794.429588	371.997558	27.259210	61.186420	18.749892	20.904609
4	2022-07-04 00:00:00	5.307278e+05	321.540266	190.816359	794.429588	371.997558	27.259210	61.186420	18.749892	20.904609
...	...	...	...	...	...	...	...	...	...	...
7008	2023-04-21 20:00:00	1.061380e+06	395.374449	224.242424	685.528121	498.737374	30.000000	68.500000	22.833333	23.916667
7009	2023-04-21 21:00:00	1.061558e+06	396.475771	225.757576	692.386831	496.527778	29.750000	69.000000	25.500000	27.250000
7010	2023-04-21 22:00:00	1.061738e+06	393.538913	227.056277	710.562414	493.686869	29.000000	69.000000	31.000000	34.833333
7011	2023-04-21 23:00:00	1.061918e+06	393.906021	225.541126	712.962963	491.161616	29.000000	69.500000	32.000000	36.750000
7012	2023-04-22 00:00:00	1.062008e+06	414.096916	220.779221	720.164609	496.212121	29.000000	70.000000	37.000000	44.000000

7013 rows × 10 columns

In [4]:
```
testl = 24*7
trainl = len(df2) - testl
df3 = df2.set_index('Datetime')

df_train = df3[:trainl]
df_test = df3[trainl-24:trainl+testl]
```

In [5]:
```
targets = ['field1','field2','field3','field4','field7','field8']
covariates = ['field5','field6']

X_train_df = df_train[covariates]
Y_train_df = df_train[targets]

X_test_df = df_test[covariates]
Y_test_df = df_test[targets]
```

In [6]:
```
X_train = TimeSeries.from_dataframe(X_train_df)
Y_train = TimeSeries.from_dataframe(Y_train_df)

X_test = TimeSeries.from_dataframe(X_test_df)
Y_test = TimeSeries.from_dataframe(Y_test_df)
```

In [7]:
```
from darts.models import XGBModel as XG
n = len(Y_test)

modelxgb1 = XG(lags=24, lags_past_covariates=24, output_chunk_length=24)
modelxgb1.fit(Y_train, past_covariates=X_train)
```

Out[7]:
```
XGBModel(lags=24, lags_past_covariates=24, lags_future_covariates=None, output_chunk_length=24, add_encoders
=None, likelihood=None, quantiles=None, random_state=None, multi_models=True, use_static_covariates=True)
```

In [8]:
```
predxgb1 = modelxgb1.predict(n, past_covariates = X_test)

predictions = TimeSeries.pd_dataframe(predxgb1)
predictions
```

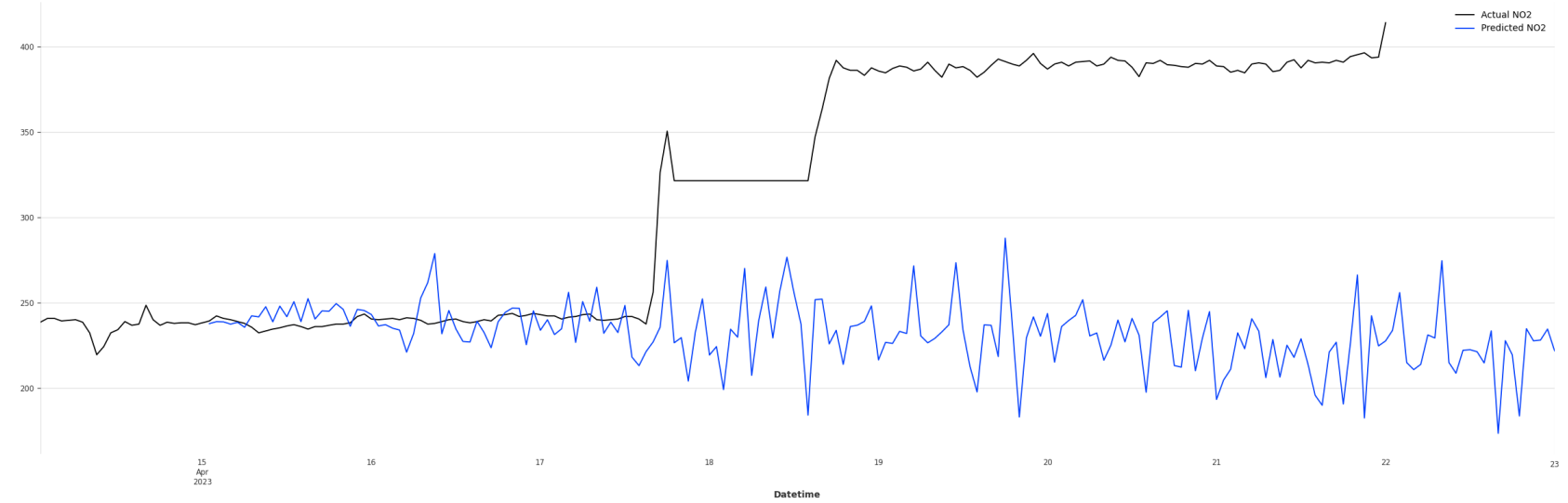
Out[8]:

component	field1	field2	field3	field4	field7	field8
Datetime						
2023-04-15 01:00:00	237.759827	132.174805	984.150879	276.548737	46.414940	56.044605
2023-04-15 02:00:00	238.948181	131.297104	897.531250	278.820984	42.577789	54.555256
2023-04-15 03:00:00	238.824020	127.378403	964.426147	260.033203	47.241184	51.383255
2023-04-15 04:00:00	237.481613	127.706001	839.864624	274.827881	45.591099	50.121376
2023-04-15 05:00:00	238.534286	128.754364	818.114441	279.527954	41.554749	51.908813
...	...	...	...	...	...	...
2023-04-22 20:00:00	234.804245	134.017258	630.830933	233.353149	7.324001	14.497888
2023-04-22 21:00:00	227.672638	125.956322	653.580139	252.825623	10.400143	15.816096
2023-04-22 22:00:00	228.121872	133.129974	681.604919	252.219223	14.059865	21.024107
2023-04-22 23:00:00	234.586533	129.786118	665.809937	258.724457	19.543341	13.162663
2023-04-23 00:00:00	221.861557	131.986954	671.453369	227.318481	19.020786	16.480936

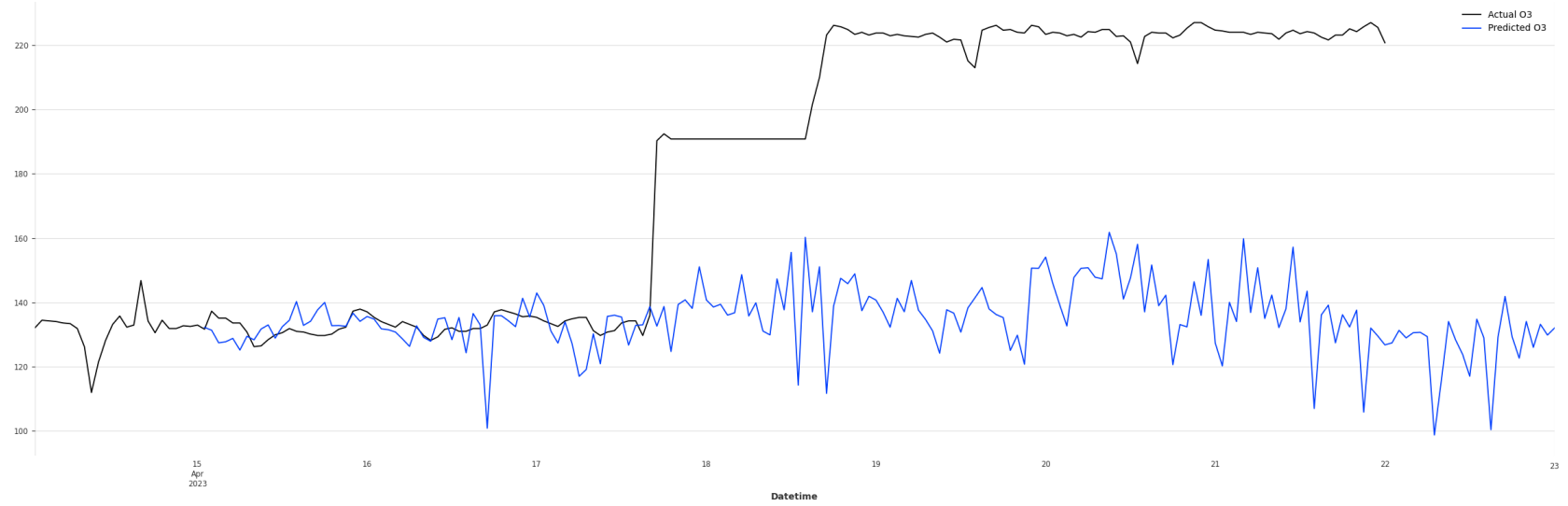
192 rows × 6 columns

In [9]:
```
from darts.metrics import rmse, mae


# Convert actual and predicted values to TimeSeries
actual_series_list = [TimeSeries.from_dataframe(Y_test_df[[target]]) for target in targets]
predicted_series_list = [TimeSeries.from_dataframe(predictions[[target]]) for target in targets]

# Calculate RMSE and MAE for each target field
rmse_values = []
mae_values = []

for actual, predicted in zip(actual_series_list, predicted_series_list):
    rmse_value = rmse(actual, predicted)
    mae_value = mae(actual, predicted)
    rmse_values.append(rmse_value)
    mae_values.append(mae_value)

# Create a DataFrame to store the results for LightGBM
results_df = pd.DataFrame({
    'Field': targets,
    'RMSE_XG': rmse_values,
    'MAE_XG': mae_values
})

# Save the LightGBM results to a CSV file
results_df.to_csv('XGBOOST_multi_results.csv', index=False)
```

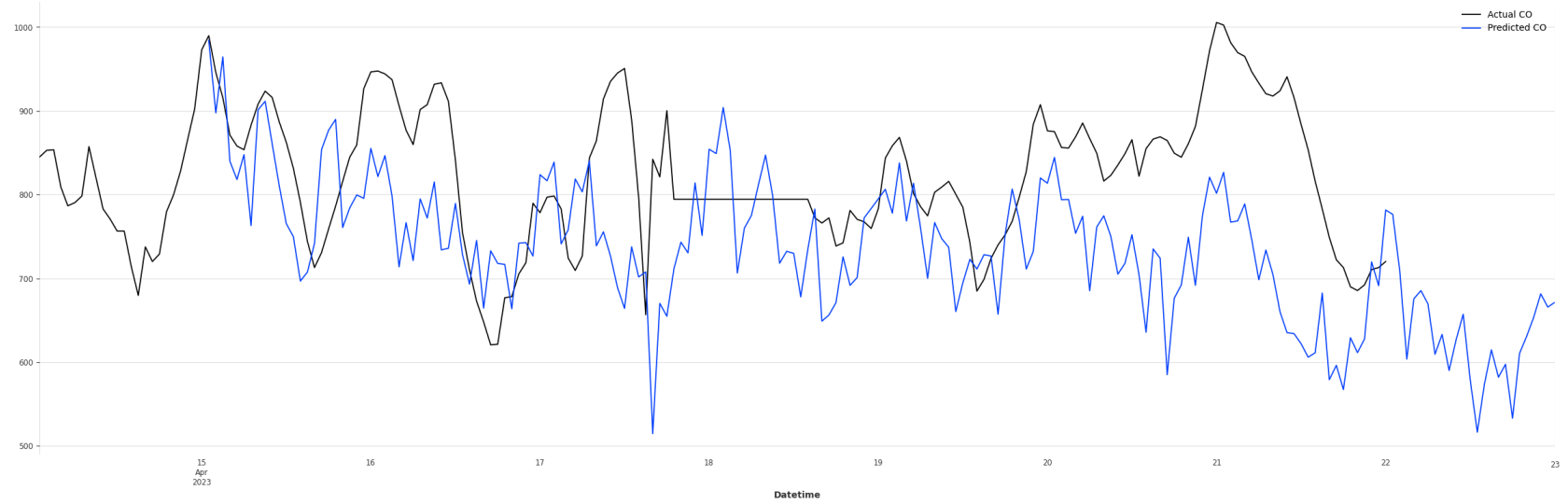
In [11]:
```
import matplotlib.pyplot as plt
plt.figure(figsize=(30,9))
Y_test_df['field1'].plot(label='Actual NO2')
predictions['field1'].plot(label='Predicted NO2')
plt.legend()
plt.show()
```



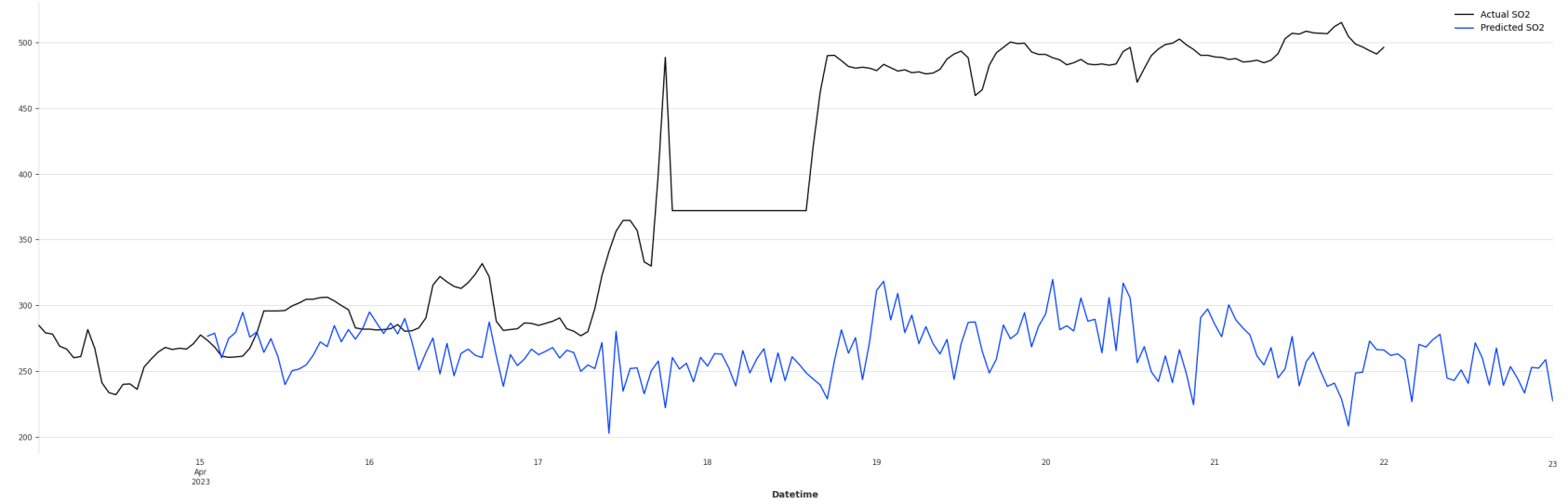
In [12]:
```
plt.figure(figsize=(30,9))
Y_test_df['field2'].plot(label='Actual O3')
predictions['field2'].plot(label='Predicted O3')
plt.legend()
plt.show()
```



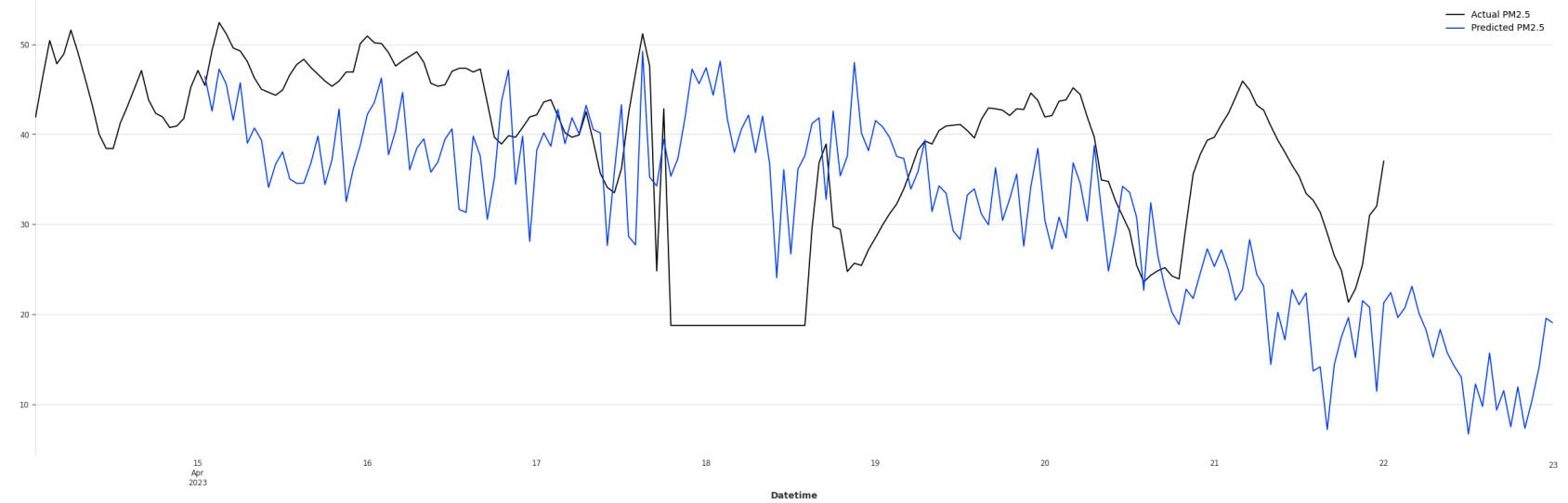
In [13]:
```
plt.figure(figsize=(30,9))
Y_test_df['field3'].plot(label='Actual CO')
predictions['field3'].plot(label='Predicted CO')
plt.legend()
plt.show()
```



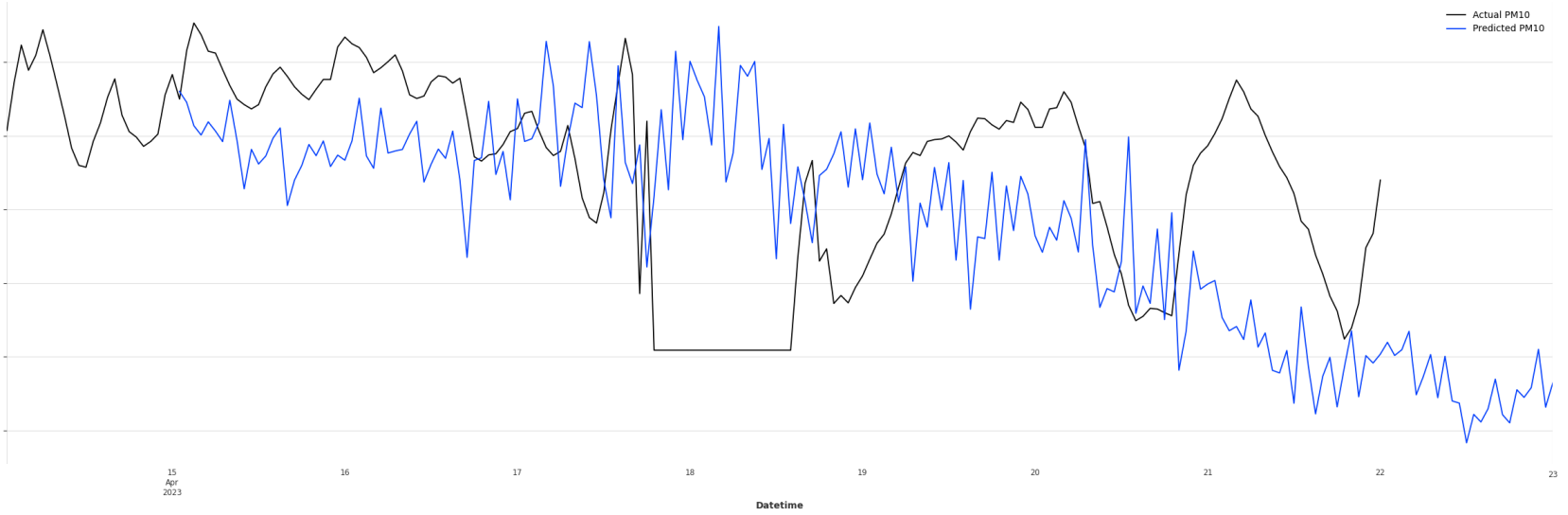
In [14]:
```
plt.figure(figsize=(30,9))
Y_test_df['field4'].plot(label='Actual SO2')
predictions['field4'].plot(label='Predicted SO2')
plt.legend()
plt.show()
```



In [15]:
```
plt.figure(figsize=(30,9))
Y_test_df['field7'].plot(label='Actual PM2.5')
predictions['field7'].plot(label='Predicted PM2.5')
plt.legend()
plt.show()
```



In [16]:
```
plt.figure(figsize=(30,9))
Y_test_df['field8'].plot(label='Actual PM10')
predictions['field8'].plot(label='Predicted PM10')
plt.legend()
plt.show()
```



In [ ]: