etl

February 25, 2026

[2]:
```
from dst_predict.imports import etl
from datetime import datetime, timezone
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import pandas as pd
```

[3]:
```
parse = etl.parse_dst_line(" Format                 IAGA-2002                ␣
  ↪                 |")
print(parse)

assert parse == "header"

parse = etl.parse_dst_line("DATE       TIME         DOY     DST              ␣
  ↪                 |")
print(parse)

assert parse == "header"

parse = etl.parse_dst_line("2000-01-01 00:00:00.000 001       -45.00")
print(parse)

expected_res = {
    "timestamp": datetime(2000, 1, 1, 0, 0, tzinfo=timezone.utc),
    "dst_nT": -45.0
}

assert parse == expected_res
```

```
header
header
{'timestamp': datetime.datetime(2000, 1, 1, 0, 0, tzinfo=datetime.timezone.utc),
'dst_nT': -45.0}
```

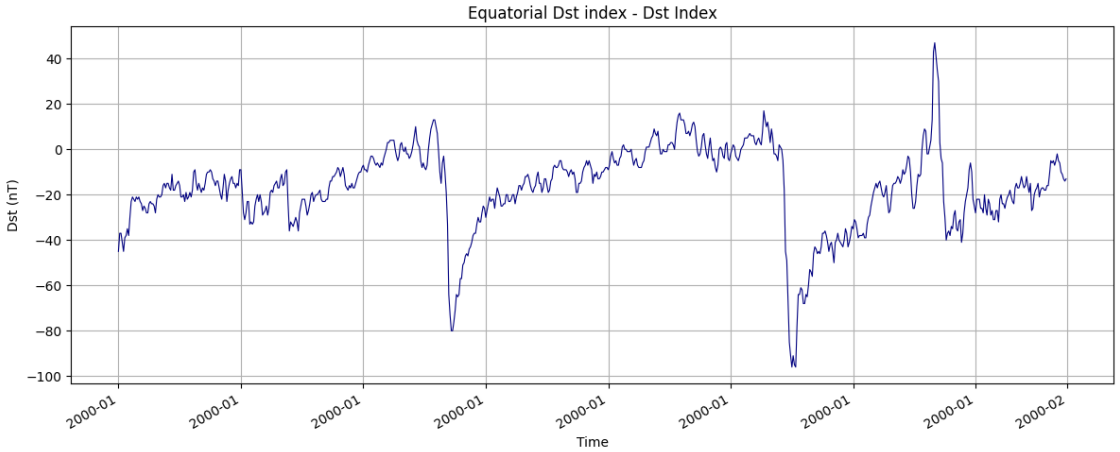
[7]:
```
file = open("../datasets/WWW_dstae01508718.dat", 'r')
records = etl.read_records(file)
rec = None
with file as f:
```

```
    rec = next(records)

print(rec["header"])
print("-" * 80)

tbl = rec["data"][0]
print(tbl["timestamp"].strftime("%Y-%m-%d %H:%M:%S"), " | " , tbl["dst_nT"])

# for tbl in rec["data"]:
#     print(tbl["timestamp"].strftime("%Y-%m-%d %H:%M:%S"), " | " ,␣
 ↪tbl["dst_nT"])
```

```
{'format': 'IAGA-2002', 'source': 'WDC for Geomagnetism, Kyoto', 'station':
'Equatorial Dst index', 'iaga_code': 'DST', 'interval': '1-hour', 'data_type':
'Final'}
--------------------------------------------------------------------------------
2000-01-01 00:00:00  |  -45.0
```

[5]:
```
def visualize_record(rec):
    header = rec.get("header", {})
    station = header.get("station", "Unknown Station")

    timestamps = [tbl["timestamp"] for tbl in rec["data"]]
    dst_values = [tbl["dst_nT"] for tbl in rec["data"]]

    plt.figure(figsize=(12, 5))
    plt.plot(timestamps, dst_values, color='navy', linewidth=0.8)

    plt.title(f"{station} - Dst Index")
    plt.xlabel("Time")
    plt.ylabel("Dst (nT)")
    plt.grid(True)

    plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%Y-%m'))
    plt.gcf().autofmt_xdate()

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

[6]:
```
visualize_record(rec)
```


Equatorial Dst index - Dst Index
Dst (nT)
40
20
0
−20
−40
−60
−80
−100
2000-01
2000-01
2000-01
2000-01
2000-01
2000-01
2000-01
2000-01
2000-02
Time