In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import seaborn as sns
sns.set(rc={'figure.figsize':(11, 4)})
from scipy import stats
from sklearn.preprocessing import StandardScaler
from itertools import chain
```

In [2]:

```python
daily_discharge = pd.read_excel("./KRB_Discharge_Stations/Discharge/Vandur_SUMQH.xls")
```

In [3]:

```python
daily_discharge.head()
```

Out[3]:

	Date	Gauge	Water Level	Discharge	Observed_Computed	Top Width	Wetted Perimeter	Radius	Velocity	Manning	Area	Slope	Max Velocity
0	1979-06-01	NaN	NaN	0.0	O	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1979-06-02	NaN	NaN	0.0	O	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	1979-06-03	NaN	NaN	0.0	O	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	1979-06-04	NaN	NaN	0.0	O	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	1979-06-05	NaN	NaN	0.0	O	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [4]:

```python
daily_discharge.isna().sum()
```

Out[4]:

```
Date                    0
Gauge                1497
Water Level          1497
Discharge               0
Observed_Computed       0
Top Width            7122
Wetted Perimeter     7122
Radius               7122
Velocity             7120
Manning              7159
Area                 7120
Slope                7158
Max Velocity         7756
Remarks             10227
dtype: int64
```

In [5]:

```python
daily_discharge.shape
```

Out[5]:

```
(10227, 14)
```

In [6]:

```python
daily_discharge = daily_discharge.set_index('Date')
```

In [7]:

```python
#xticks = pd.date_range(start="1980-01-01",end="2004-12-31", freq='Y')
#daily_discharge['Discharge'].plot(linewidth=1.0, ylim=(0,2.5), xticks=xticks);
daily_discharge.loc['1966-12':,'Discharge'].plot(marker='.', alpha=0.5, linestyle='None',
figsize=(11, 9), subplots=True)
```

Out[7]:

```
array([<AxesSubplot:xlabel='Date'>], dtype=object)
```


In [8]:
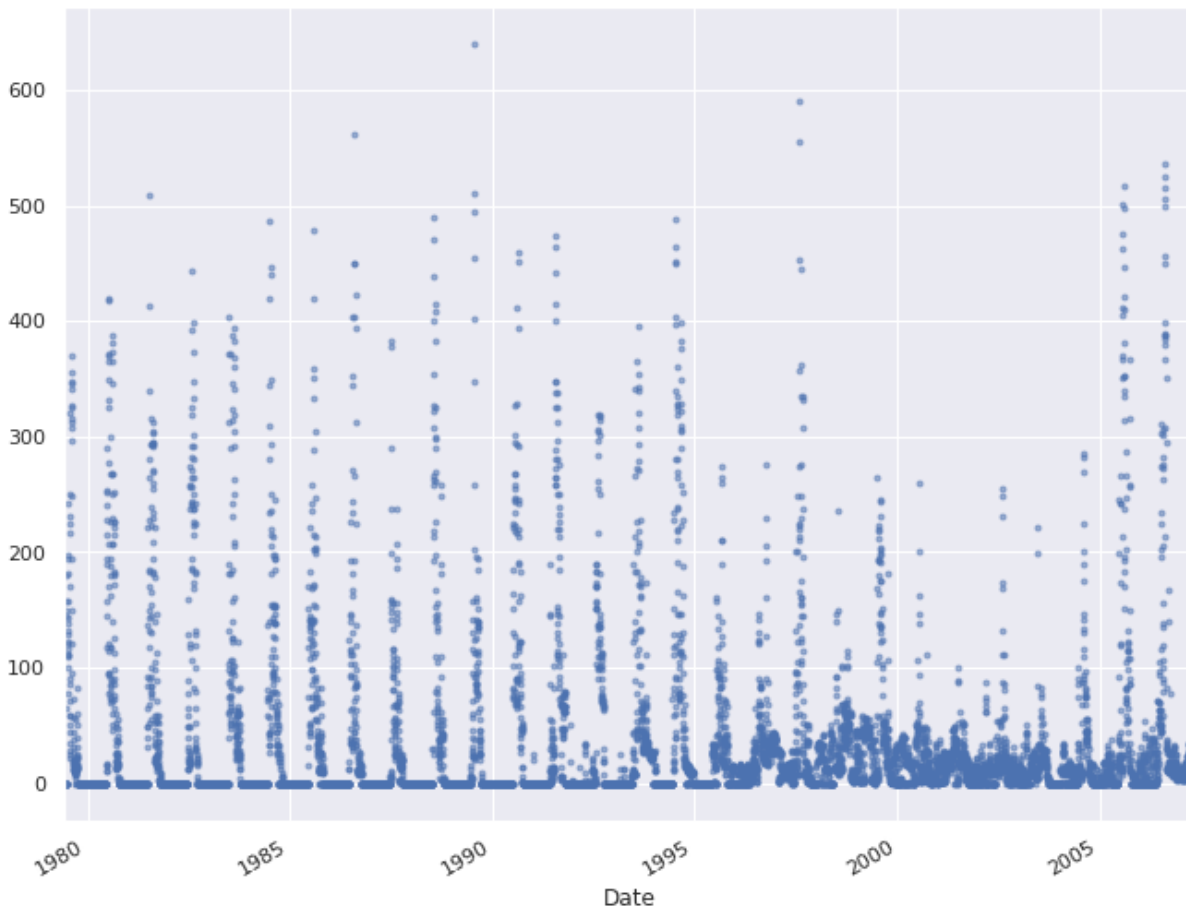
```python
years = mdates.YearLocator()    # every year
months = mdates.MonthLocator()  # every month
years_fmt = mdates.DateFormatter('%Y')

fig, ax = plt.subplots(figsize=(11, 9))
ax.plot(daily_discharge.loc['1967-01':, 'Discharge'])
ax.set_ylabel('Discharge')
# Set x-axis major ticks to weekly interval, on Mondays
datemin = np.datetime64(daily_discharge.index[0], 'Y')
datemax = np.datetime64(daily_discharge.index[5000], 'Y') + np.timedelta64(1, 'Y')
#ax.set_xlim(datemin, datemax)
#ax.set_ylim(0, 2.5)
ax.xaxis.set_major_locator(years)
ax.xaxis.set_major_formatter(years_fmt)
ax.xaxis.set_minor_locator(months)

ax.format_xdata = mdates.DateFormatter('%Y-%m-%d')
ax.format_ydata = lambda x: '$%1.2f' % x   # format the price.
ax.grid(True)
fig.autofmt_xdate()
```
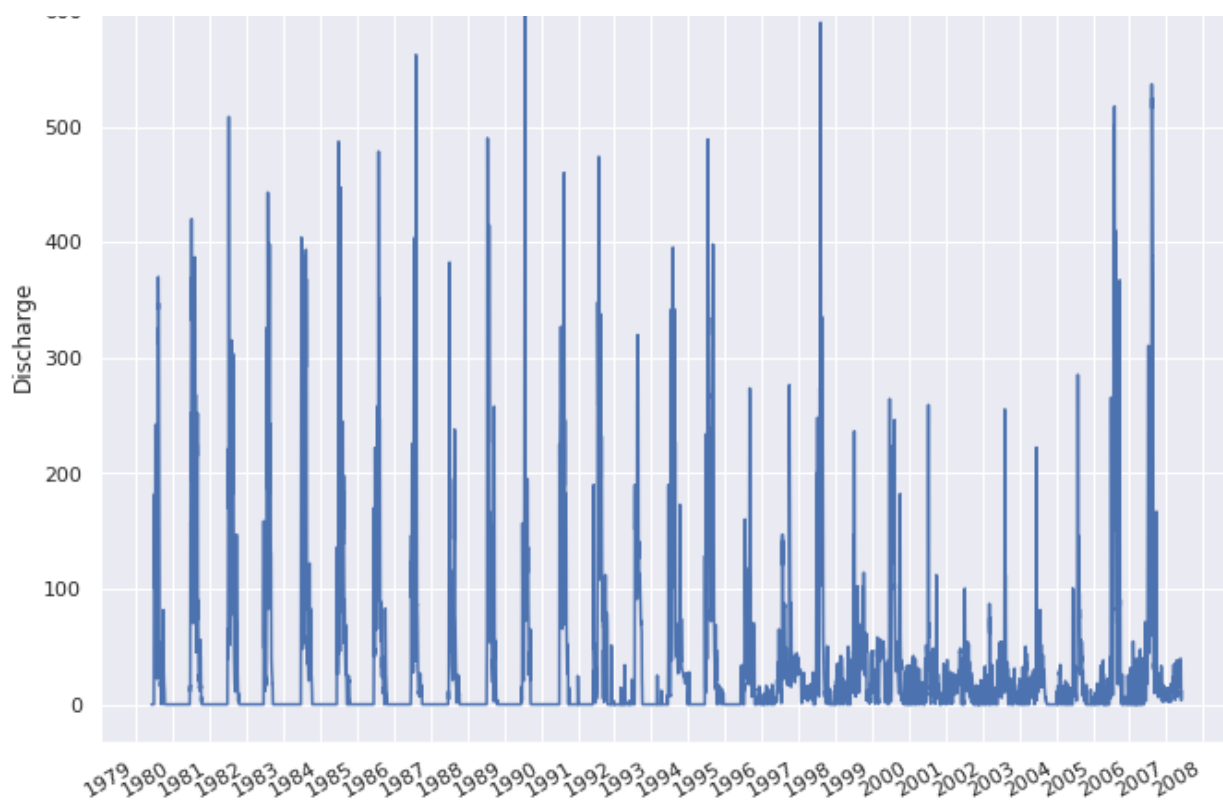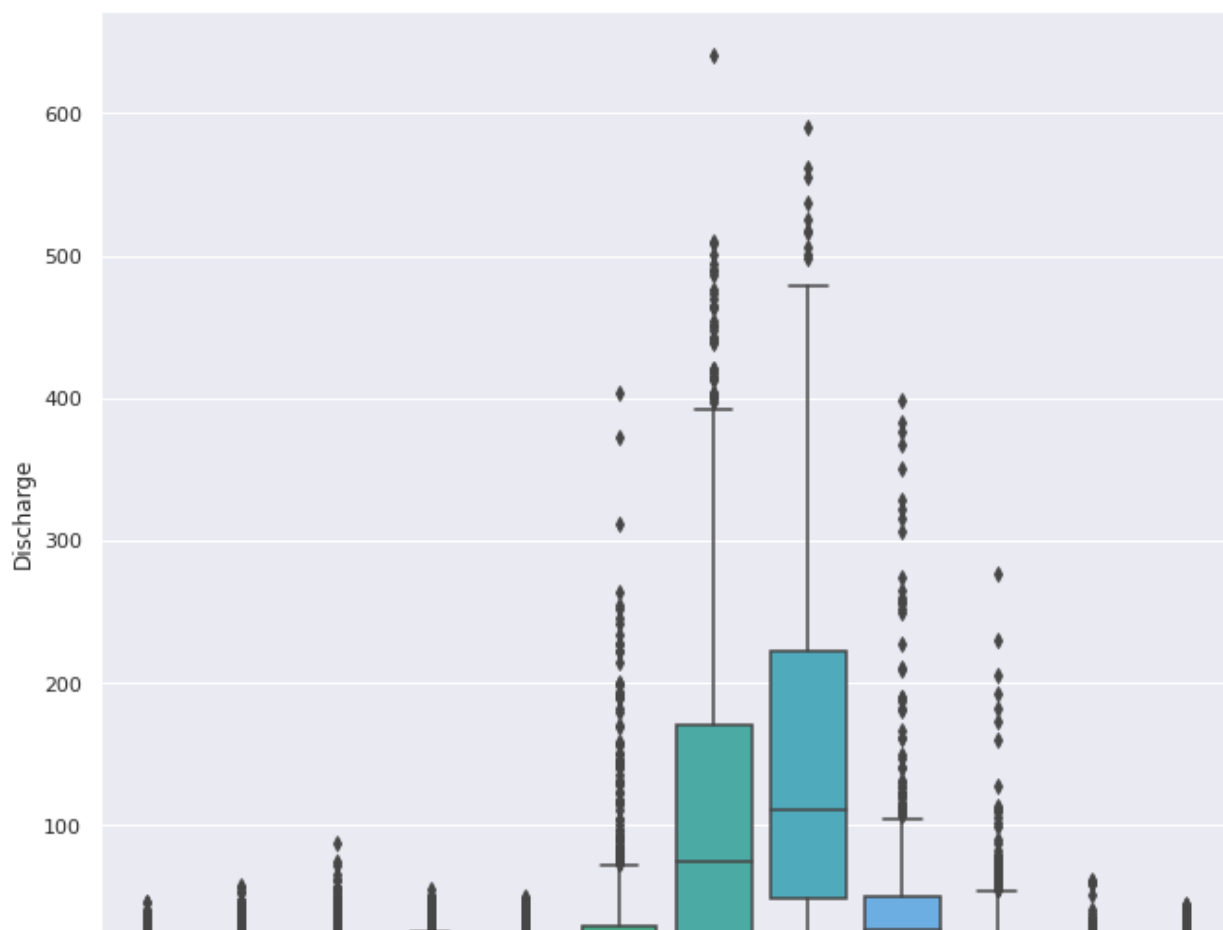
```python
daily_discharge['Year'] = daily_discharge.index.year
daily_discharge['Month'] = daily_discharge.index.month
```
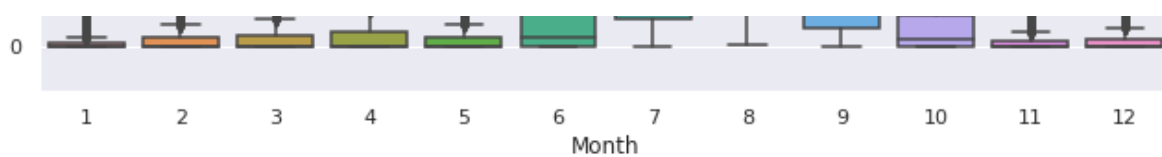
```python
fig, axes = plt.subplots(1, 1, figsize=(11, 10), sharex=True)
sns.boxplot(data=daily_discharge, x='Month', y='Discharge', ax=axes)
ax.set_title('Discharge')
```

```
Text(0.5, 1.0, 'Discharge')
```

```
# Specify the data columns we want to include (i.e. exclude Year, Month, Weekday Name)
data_columns = ['Gauge', 'Water Level', 'Discharge', 'Observed_Computed', 'Remarks']
# Resample to weekly frequency, aggregating with mean
discharge_weekly_mean = daily_discharge[data_columns].resample('W').mean()
discharge_monthly_mean = daily_discharge[data_columns].resample('M').mean()
```

Resampling

In [12]:

```
fig, ax = plt.subplots()
fig.set_size_inches(20,8)
ax.plot(discharge_monthly_mean['Discharge'], color='black', label='Discharge')
#discharge_monthly['Discharge'].plot.area(ax=ax, linewidth=0)
ax.xaxis.set_major_locator(mdates.YearLocator())
ax.xaxis.set_major_formatter(mdates.DateFormatter('%Y'))
ax.xaxis.set_minor_locator(mdates.MonthLocator())
ax.legend()
ax.set_ylabel('Monthly Total');
ax.format_xdata = mdates.DateFormatter('%Y-%m-%d')
ax.format_ydata = lambda x: '$%1.2f' % x   # format the price.
ax.grid(True)
fig.autofmt_xdate()
```
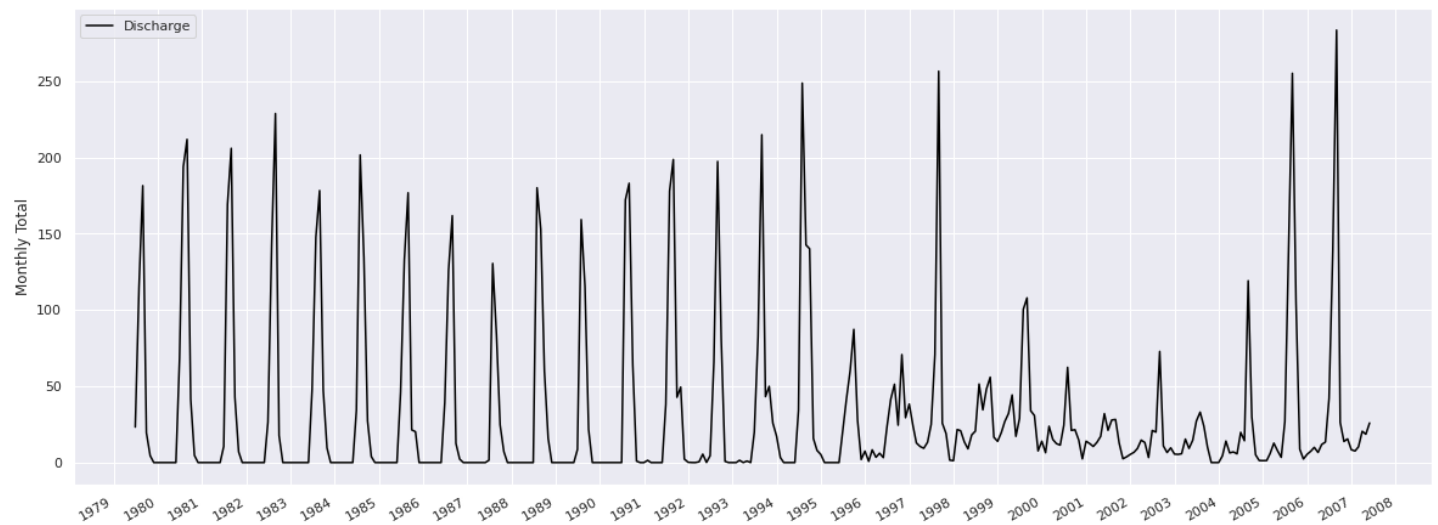

Rolling windows

Weekly and yearly rolling window

In [13]:

```
discharge_7d = daily_discharge[data_columns].rolling(7, center=True).mean()
discharge_30d = daily_discharge[data_columns].rolling('30D').mean()
discharge_365d = daily_discharge[data_columns].rolling(window=365, center=True, min_peri
ods=360).mean()
```

In [14]:

```
# Start and end of the date range to extract
start, end = '1980-01','1990-02'
# Plot daily, weekly resampled, and 7-day rolling mean time series together
fig, ax = plt.subplots()
fig.set_size_inches(20,8)
```

```
#ax.plot(daily_discharge.loc[start:end, 'Discharge'],
#marker='.', linestyle='-', linewidth=0.5, label='Daily')
ax.plot(discharge_weekly_mean.loc[start:, 'Discharge'],
marker='o', markersize=8, linestyle='-', linewidth=0.1, label='Weekly Mean Resample')
ax.plot(discharge_7d.loc[start:, 'Discharge'],
marker='.', linestyle='-', linewidth=0.5, label='7-d Rolling Mean')
ax.set_ylabel('Discharge)')
ax.legend();
```
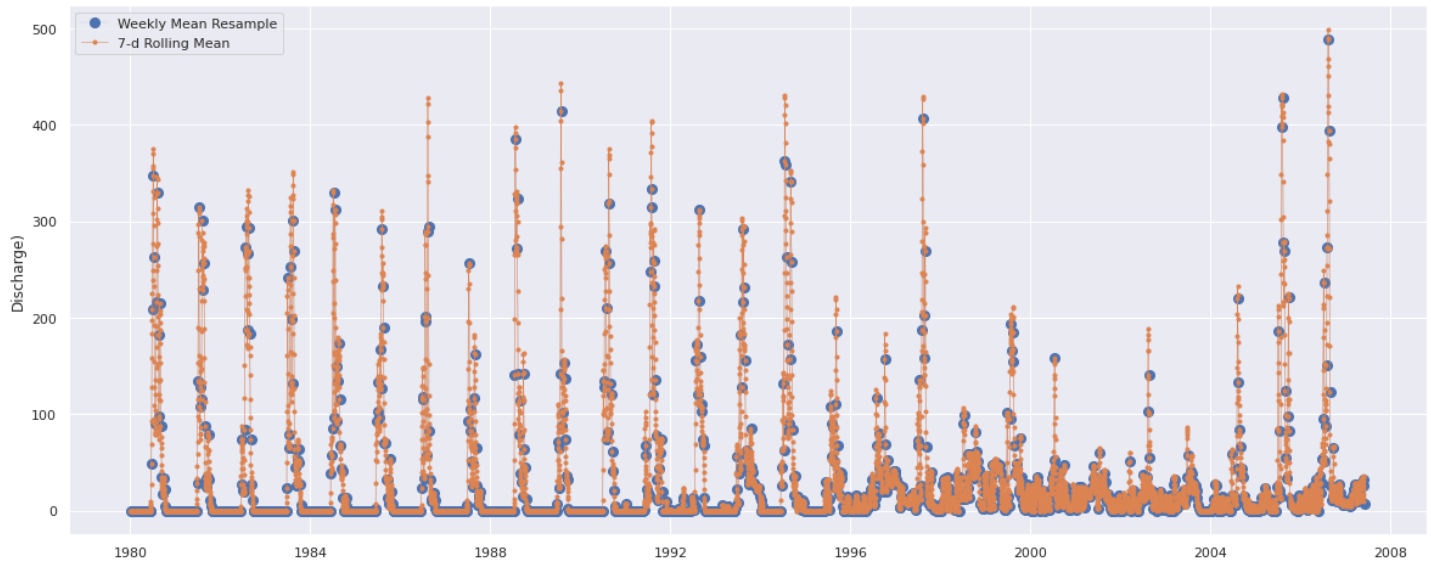

In [15]:

```
# Plot daily, 7-day rolling mean, and 365-day rolling mean time series
fig, ax = plt.subplots()
fig.set_size_inches(20,8)
#ax.plot(daily_discharge['Discharge'], marker='.', markersize=2, color='0.6',
#linestyle='None', label='Daily')
#ax.plot(discharge_7d['Discharge'], linewidth=2, label='7-d Rolling Mean')
ax.plot((discharge_365d['Discharge']-999)*0.001, color='0.2', linewidth=3,
label='Trend (365-d Rolling Mean)')
# Set x-ticks to yearly interval and add legend and labels
ax.xaxis.set_major_locator(mdates.YearLocator())
ax.xaxis.set_major_formatter(mdates.DateFormatter('%Y'))
ax.xaxis.set_minor_locator(mdates.MonthLocator())
ax.legend()
ax.set_xlabel('Year')
ax.set_ylabel('Discharge')
ax.format_xdata = mdates.DateFormatter('%Y-%m-%d')
ax.format_ydata = lambda x: '$%1.2f' % x   # format the price.
ax.grid(True)
fig.autofmt_xdate()
```
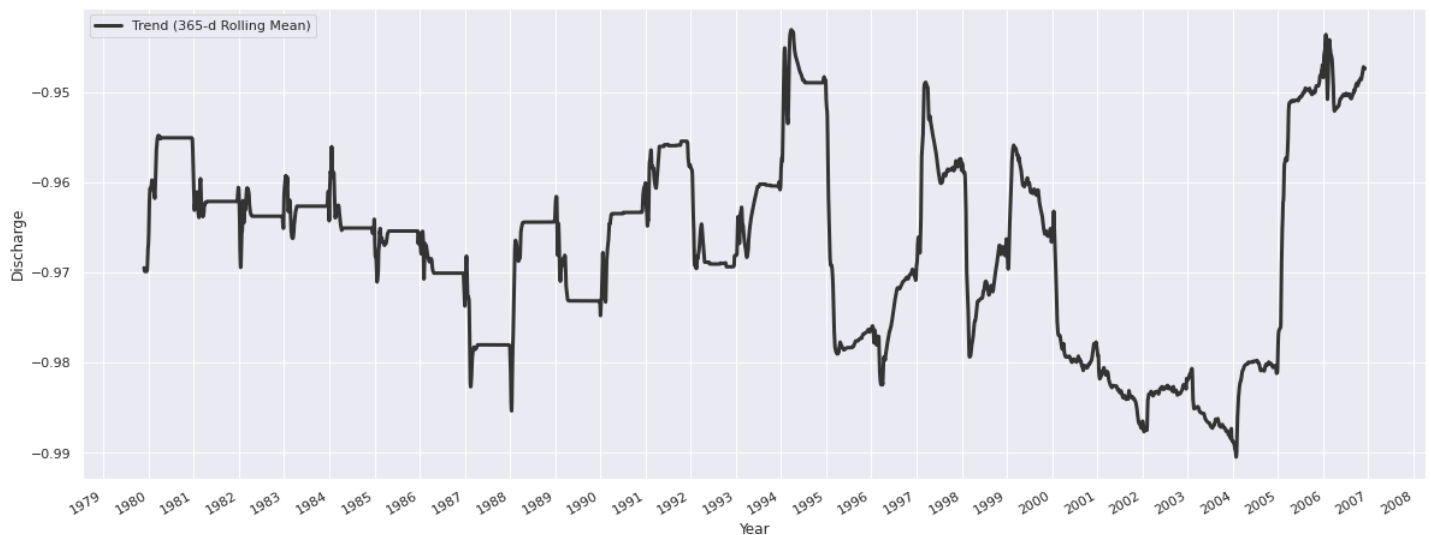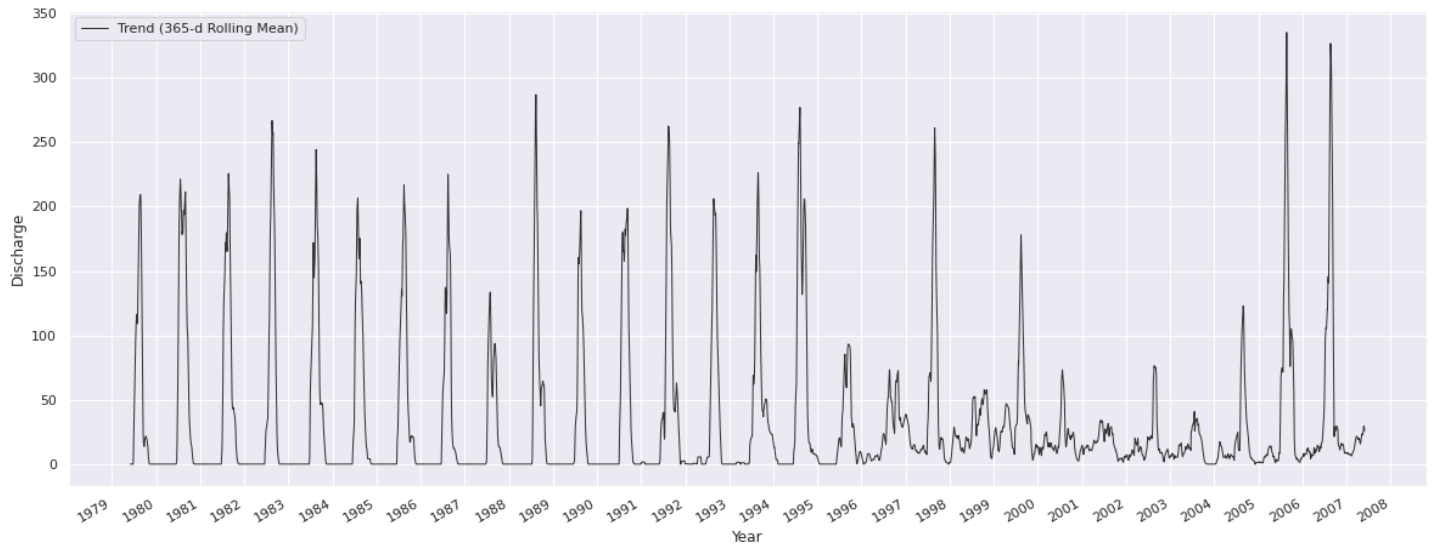

In [16]:

```
# Plot daily, 7-day rolling mean, and 365-day rolling mean time series
```

```
fig, ax = plt.subplots()
fig.set_size_inches(20,8)
#ax.plot(daily_discharge['Discharge'], marker='.', markersize=2, color='0.6',
#linestyle='None', label='Daily')
#ax.plot(discharge_7d['Discharge'], linewidth=2, label='7-d Rolling Mean')
ax.plot(discharge_30d['Discharge'], color='0.2', linewidth=1,
label='Trend (365-d Rolling Mean)')
# Set x-ticks to yearly interval and add legend and labels
ax.xaxis.set_major_locator(mdates.YearLocator())
ax.xaxis.set_major_formatter(mdates.DateFormatter('%Y'))
ax.xaxis.set_minor_locator(mdates.MonthLocator())
ax.legend()
ax.set_xlabel('Year')
ax.set_ylabel('Discharge')
ax.format_xdata = mdates.DateFormatter('%Y-%m-%d')
ax.format_ydata = lambda x: '$%1.2f' % x   # format the price.
ax.grid(True)
fig.autofmt_xdate()
```


In [ ]: