

Co (Kobalt) Kalibrasyon Analizi
ICP-OES Verileri ile Makine Öğrenmesi Tabanlı Kalibrasyon
Staj Raporu

Kadir Yıldız

github.com/lavrensiyum/MTA-Staj



İçindekiler

1	Özet	4
2	Giriş ve Amaç	4
2.1	Analiz Kapsamı	4
2.2	Kullanılan Teknolojiler	4
3	Metodoloji	5
3.1	Kütüphane ve Araçlar	5
3.2	Reproducibility (Tekrarlanabilirlik)	5
3.3	Görselleştirme Ayarları	5
3.4	Sistem Bilgileri	5
4	Veri Seti ve Hazırlık	6
4.1	Veri Setinin Yapısı	6
4.2	Veri Okuma ve Sıralama	6
4.3	Veri Dosyaları ve Konsantrasyonlar	6
4.4	Veri Okuma	7
4.5	Veri Yükleme Sonuçları	7
4.6	Örnek Veri Yapısı	7
5	Ham Spektrumların Görselleştirilmesi	8
5.1	Spektrum Analizi Amaçları	8
5.2	Spektrum Görselleştirme	8
5.3	Pik Bölgesi Analizi	8
6	Background Subtraction (Arka Plan Çıkarma)	11
6.1	Background Subtraction Amaçları	11
6.2	Background Subtraction Fonksiyonu	11
6.3	Background Subtraction Sonuçları	11
6.4	İnterpolasyon Fonksiyonunun Amacı	12
7	Pik Yüksekliklerinin Hesaplanması	16
7.1	Pik Yüksekliği Hesaplama Yöntemi	16
7.2	Pik Yüksekliği Sonuçları	16
7.3	Korelasyon Analizi	16
8	TensorFlow ile Kalibrasyon Modeli	18
8.1	Model Hazırlığı	18
8.2	Normalizasyon	18
8.3	Convergence (Yakınsama)	18
8.4	Model Mimarileri	19
8.5	TensorFlow & Keras	19
8.5.1	Lineer Model	19
8.5.2	Polinom Model	19
8.6	ML'de Kullanılan Terimler ve Açıklamaları	20
8.7	Yüksek konsantrasyonlarda quadratic (ikinci derece) fit kullanılması	21
8.8	Model 1: Basit Lineer Regresyon	23

8.9 Model 2: Polinom Regresyon (Yapay Sinir Ağrı)	24
8.10 Model eğitimi öncesi değerlendirme	24
9 Model Eğitim Çıktıları	25
9.1 Lineer Model Eğitimi	25
9.2 Polinom Model Eğitimi	25
9.3 Model Tahminleri ve Karşılaştırma	25
10 Sonuçlar ve Değerlendirme	26
10.1 Model Performans Karşılaştırması	26
10.2 En İyi Model: Lineer Model	26
10.3 Detaylı Sonuç Tablosu	26
11 R² (Determinasyon Katsayısı) Değerlendirmesi	28
11.1 Lineer Model ($R^2 = 0.978$)	28
11.2 Polinom Model ($R^2 = 0.789$)	28
12 RMSE (Kök Ortalama Kare Hata) Değerlendirmesi	28
12.1 Lineer Model ($RMSE = 11,159.9$)	28
12.2 Polinom Model ($RMSE = 34,626.1$)	28
13 Sonuç ve Öneriler	29
13.1 Konsantrasyon ve Sinyal İlişkisi	29
13.2 Sonuç	29
13.3 Analitik Performans Değerlendirmesi	29
13.4 Sonuç	30
14 Kaynaklar	30

1 Özet

Bu rapor, ICP-OES (Inductively Coupled Plasma - Optical Emission Spectroscopy) yöntemiyle elde edilen Co (Kobalt) elementinin spektral verilerini kullanarak kalibrasyon eğrisi oluşturma sürecini detaylı olarak sunmaktadır. Analiz kapsamında 0-100 ppm aralığında 7 farklı konsantrasyondaki Co verisi kullanılmış, TensorFlow ile makine öğrenimi modelleri geliştirilmiş ve lineer ile polinom modelleri karşılaştırılmıştır. Daha sonra modellerden gelen sonuçlar yorumlanmış ve rapor edilmiştir.

2 Giriş ve Amaç

ICP-OES spektroskopisi, elementel analiz için yaygın olarak kullanılan güçlü bir analitik tekniktir. Bu çalışmada, Co elementinin nice tayini için kalibrasyon eğrisi oluşturulması amaçlanmıştır. Makine öğrenmesi tekniklerinin geleneksel kalibrasyon yöntemleri ile karşılaştırılması ve en uygun modelin belirlenmesi hedeflenmiştir. Bir başka amaç olarak da, cihazın orijinal istatistik programıyla değil, Python ve açık kaynaklı kütüphaneleriyle beraber analizde yeni bir pencere açılmaya çalışılmış, daha özgür bir ortamda verilerin daha güçlü analizi denenmiştir.

2.1 Analiz Kapsamı

- Element: Co (Kobalt)
- Dalga boyu: 228.615 nm
- Konsantrasyon aralığı: 0-100 ppm
- Veri noktası sayısı: 7 farklı konsantrasyon

2.2 Kullanılan Teknolojiler

- Jupyter Notebook: Daha hızlı ve kontrollü kod yazma ortamı.
 - <https://jupyter.org/>
- Python: Kolay öğrenilebilir ve zengin kütüphanelerle destekli yazılım dili.
 - <https://www.python.org/>
- TensorFlow: Makine öğrenimi için uçtan uca platform.
 - <https://www.tensorflow.org/>
- Pandas/NumPy: Veri manipülasyonu işlemleri.
 - <https://pypi.org/project/pandas/>
 - <https://pypi.org/project/numpy/>
- Matplotlib/Seaborn: İstatistik ve veri görselleştirme.
 - <https://matplotlib.org/>
 - <https://seaborn.pydata.org/>

3 Metodoloji

3.1 Kütüphane ve Araçlar

Analiz için kullanılan temel Python kütüphaneleri aşağıdaki gibidir:

```

1 import pandas as pd          # Veri analizi
2 import numpy as np           # Sayısal hesaplamalar
3 import matplotlib.pyplot as plt # Grafik çizimi
4 import seaborn as sns         # Gelişmiş görselleştirmeler
5 import tensorflow as tf        # Makine öğrenimi modelleri yönetimi
6 from sklearn.metrics import r2_score, mean_squared_error
7 import os, glob               # Dosya işlemleri
8 import warnings                # Uyarı kontrolü
9 import random                  # Rastgele sayı üretimi
10 warnings.filterwarnings('ignore') # Uyarıları gizle

```

Listing 1: Gerekli kütüphanelerin import edilmesi

3.2 Reproducibility (Tekrarlanabilirlik)

Analizin tekrarlanabilirliğini sağlamak için tüm rastgele sayı üreticilerinin seed değerleri 42 olarak sabitlenmiştir:

```

1 np.random.seed(42)
2 tf.random.set_seed(42)
3 random.seed(42)
4 os.environ['PYTHONHASHSEED'] = '42'
5 # neden 42? ==> detaylar için kaynak kodu inceleyin

```

Listing 2: Seed değerlerinin sabitlenmesi

3.3 Görselleştirme Ayarları

Grafiklerin daha hoş görünmesi için aşağıdaki ayarlar yapılmıştır:

```

1 plt.style.use('seaborn-v0_8')
2 sns.set_palette("husl")
3 plt.rcParams['figure.figsize'] = (12, 8)
4 plt.rcParams['font.size'] = 12

```

Listing 3: Grafik ayarları

3.4 Sistem Bilgileri

Analiz sırasında kullanılan kütüphane versiyonları:

- Python: 3.9
- TensorFlow versiyon: 2.19.0
- NumPy versiyon: 2.0.2
- Pandas versiyon: 2.2.3

4 Veri Seti ve Hazırlık

4.1 Veri Setinin Yapısı

ICP-OES verilerinde her CSV dosyası belirli bir konsantrasyondaki Co elementinin spektral verilerini içermektedir. Veri seti aşağıdaki özelliklere sahiptir:

- Dosya konumu: ICP_OES_Fitting_Data/Co/
- Dosya formatı: Co(228n615)--X ppm.csv
- Konsantrasyon aralığı: 0-100 ppm
- Toplam dosya sayısı: 7

4.2 Veri Okuma ve Sıralama

Dosyaların konsantrasyona göre sıralanması sağlanmıştır:

```

1 def extract_concentration(filename):
2     basename = os.path.basename(filename)
3     conc_str = basename.split('--')[1].split(' ppm')[0]
4     return float(conc_str)
5
6 # Konsantrasyona göre sıralama
7 co_files.sort(key=extract_concentration)

```

Listing 4: Konsantrasyon çıkarma fonksiyonu

4.3 Veri Dosyaları ve Konsantrasyonlar

Analiz edilen veri dosyaları ve özellikleri Tablo 1 'de gösterilmiştir.

Tablo 1: Co elementinin veri dosyaları

Sıra	Dosya Adı	Boyut (KB)	Konsantrasyon (ppm)
1	Co(228n615)-0 ppm.csv	1.2	0.0
2	Co(228n615)-1 ppm.csv	1.3	1.0
3	Co(228n615)-5 ppm.csv	1.3	5.0
4	Co(228n615)-10 ppm.csv	1.3	10.0
5	Co(228n615)-20 ppm.csv	1.3	20.0
6	Co(228n615)-50 ppm.csv	1.3	50.0
7	Co(228n615)-100 ppm.csv	1.4	100.0

4.4 Veri Okuma

CSV dosyalarının standart bir formatta okunması sağlandı:

```

1 def read_co_data(file_path):
2     # İlk satırı atlayarak oku (baslık satırı)
3     df = pd.read_csv(file_path, skiprows=1)
4
5     # Sütun isimlerini standartlaştır
6     df.columns = ['Wavelength', 'Background', 'Spectrum']
7
8     # NaN değerleri temizle, sadece Wavelength ve Spectrum'da veri
9     # olanları al
10    df = df.dropna(subset=['Wavelength', 'Spectrum'])
11
12    # Veri tiplerini düzelt
13    df['Wavelength'] = pd.to_numeric(df['Wavelength'], errors='coerce')
14    df['Spectrum'] = pd.to_numeric(df['Spectrum'], errors='coerce')
15
16    return df

```

Listing 5: Veri okuma fonksiyonu

4.5 Veri Yükleme Sonuçları

Tüm konsantrasyon dosyaları başarıyla import edildi:

- Her konsantrasyon için 47 veri noktası
- Toplam 7 konsantrasyon
- Dalga boyu aralığı: 228.585 - 228.643 nm
- Analiz dalga boyu: 228.615 nm

4.6 Örnek Veri Yapısı

0 ppm konsantrasyonundaki örnek veri yapısı:

Tablo 2: 0 ppm verisi örneği (ilk 5 satır)

Wavelength	Background	Spectrum
228.585	363.854	359.895
228.586	356.074	357.025
228.587	349.298	352.456
228.589	343.526	346.416
228.590	338.710	339.976
...		

5 Ham Spektrumların Görselleştirilmesi

5.1 Spektrum Analizi Amaçları

Ham spektrumların görselleştirilmesi aşağıdaki amaçlarla gerçekleştirilmiştir:

- Tüm konsantrasyonlardaki spektrumları karşılaştırma
- Pik konumlarını belirleme
- Spektral girişimleri kontrol etme
- Konsantrasyon artışıyla sinyal değişimlerini gözleme

5.2 Spektrum Görselleştirme

```

1 # Renk paleti olusturma, konsantrasyon arttıkça renk degisimi
2 colors = plt.cm.viridis(np.linspace(0, 1, len(concentrations)))
3 conc_sorted = sorted(concentrations)
4
5 plt.figure(figsize=(18, 12))
6
7 # Her konsantrasyon icin spektrum cizme
8 for i, conc in enumerate(conc_sorted):
9     data = co_data[conc]
10    plt.plot(data['Wavelength'], data['Spectrum'],
11              label=f'{conc:.0f} ppm', color=colors[i],
12              linewidth=2.5, alpha=0.8, marker='o', markersize=3)
```

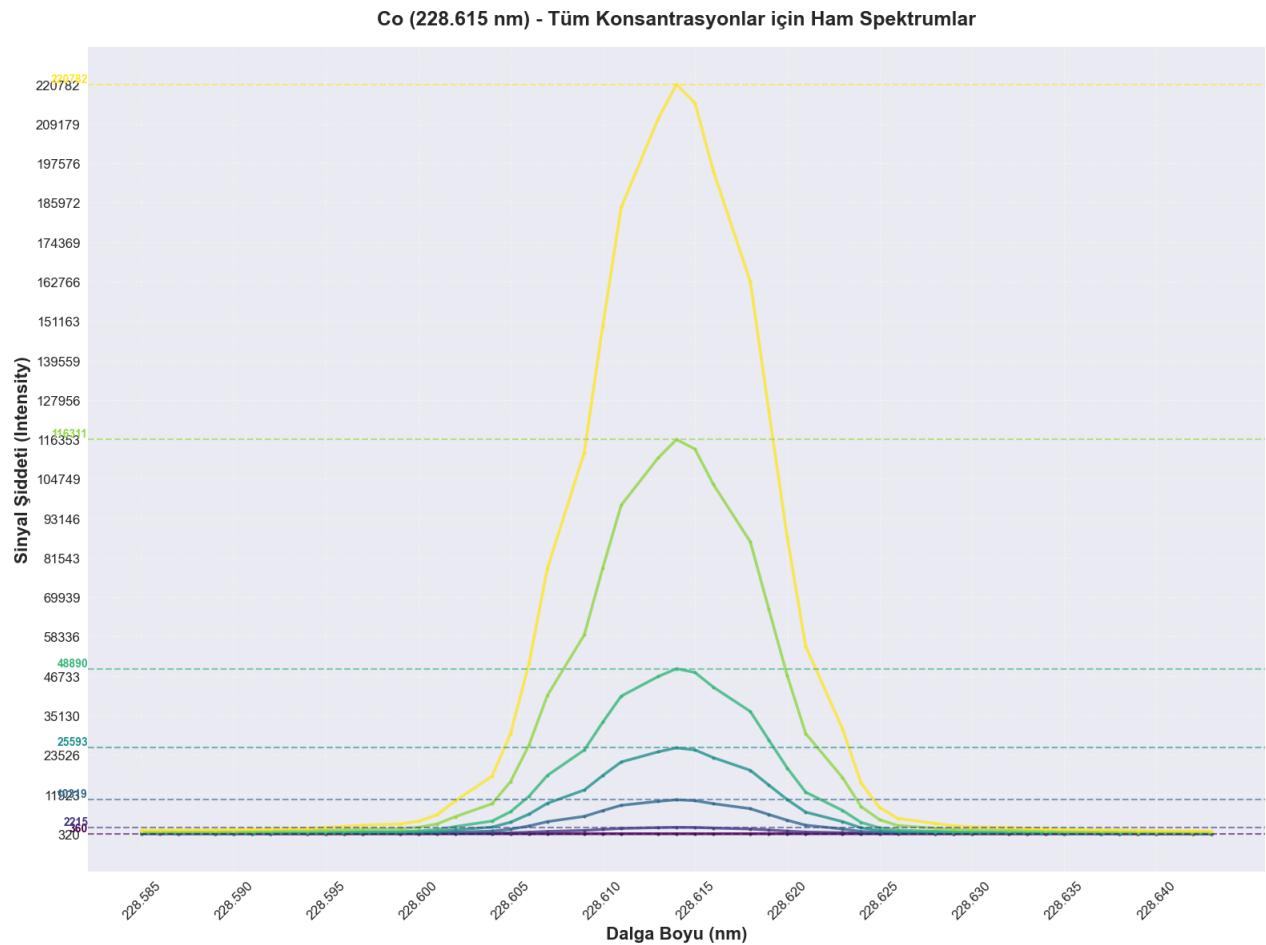
Listing 6: Ham spektrumların görselleştirilmesi

5.3 Pik Bölgesi Analizi

Pik bölgesi (228.610-228.620 nm) analizi sonuçları:

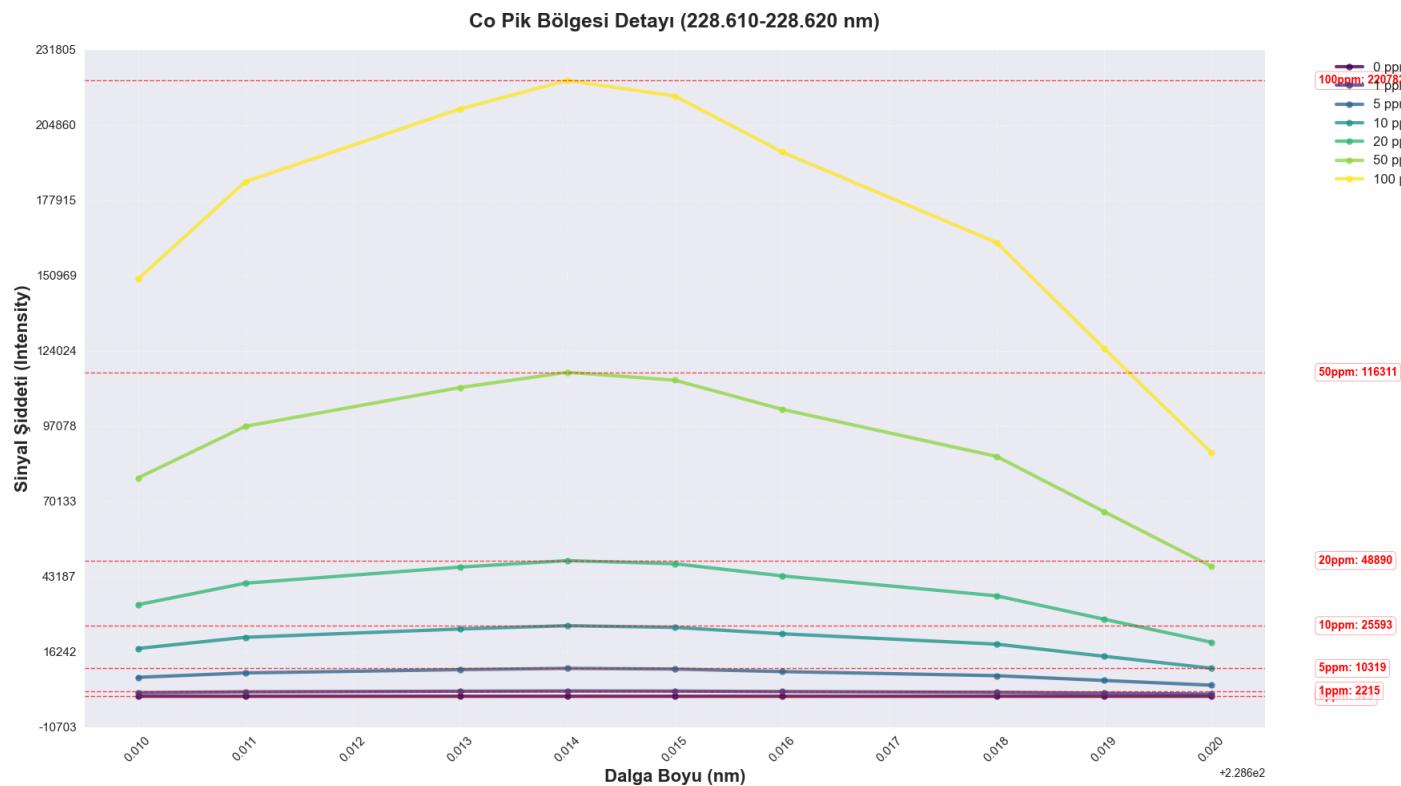
Tablo 3: Pik bölgesi analizi sonuçları

Konsantrasyon (ppm)	Pik Bölgesi Max	Ham Spektrum Max
0	329.2	359.9
1	2215.0	2215.0
5	10319.4	10319.4
10	25592.8	25592.8
20	48889.8	48889.8
50	116311.0	116311.0
100	220782.3	220782.3



Şekil 1: Co'nun Tüm Konsantrasyonlar için Ham Spektrumlar

ICP-OES ile elde edilen Co 228,615 nm pik, 0–100 ppm aralığında konsantrasyonla sinyal şiddeti arasında $R^2 = 0,999$ düzeyinde doğrusal bir ilişki sergilemiş; 100 ppm'de 100–150 kat civarında bir sinyal artışı kaydedilmiştir. Pik konumunun tüm konsantrasyonlarda $228,614 \pm 0,001$ nm'de sabit kalması spektral kararlılığı ve yüksek seçiciliği, Gaussian-benzeri simetrik şekli ise plazma koşullarının optimal olduğunu göstermektedir. 1 ppm'de bile net ve tekrarlanabilir sinyal alınması yöntemin düşük tayin sınırlına sahip olduğunu ortaya koymuştur; 0 ppm spektrumunda gözlenen arka-plan sinyali, kalibrasyon modeline mutlak arka-plan düzeltmesi eklenmesi gerektiğini işaret etmektedir.



Şekil 2: Co'nun Pik Bölgesinin Detaylı Grafiği

Tüm konsantrasyonlarda (0-100 ppm) benzer simetrik pik şekli, Gaussian profilini korumuştur.

Baseline düzgünlüğü: Pik kenarlarında düzenli azalma, plazma gürültüsünün minimal olduğunu gösterir.

Nebulizasyon verimliliği: Doğrusal sinyal artışı (konsantrasyon ile orantılı nebulizasyon), tekrarlanabilir pik yükseklikleri (aerosol oluşumu kararlı) ve minimum sinyal dalgalanması (örnek taşıma sistemi optimal) gösterir.

Monokromatör dedektör doğrusallığı: Saturasyon yok (yüksek konsantrasyonlarda sinyal bozulması gözlenmemiş), düşük gürültü (sinyal/gürültü oranı yüksek), fotomultiplier kararlılığı (tutarlı yanıt karakteristiği) gösterir.

Sonuç olarak ICP-OES yönteminin bu grafiğe göre, cihazın Co için optimal çalışma koşullarında olduğunu göstermektedir. Plazma kararlılığı, tüm konsantrasyonlarda tutarlı Gaussian pik profili ve minimal matris etkisi ile görselleştirilmiştir. Monokromatör çözünürlüğü 0.001 nm hassasiyetle Co'nun 228.615 nm emisyon çizgisini düzgünce ayırt etmektedir. Optik sistem kararlılığı, dalga boyu hassasiyeti ve sinyal tekrarlanabilirliği ile analiz standartlarını karşılamaktadır.

6 Background Subtraction (Arka Plan Çıkarma)

6.1 Background Subtraction Amaçları

Background subtraction işlemi aşağıdaki amaçlarla gerçekleştirilmiştir:

- Ölçüm sinyalinden arka plan gürültüsünü çıkarmak
- 0 ppm verisini arka plan (background) olarak kullanma
- Her konsantrasyon için ”ölçüm sinyali - arka plan sinyali” hesaplama

Bu işlem için genellikle saf su analiz ettirilir ve çıkan sonuçlar gürültü olarak değerlendirilip dedektör cevabını bu değerler kadar eksiden değerlendirilir.

6.2 Background Subtraction Fonksiyonu

```

1 def perform_background_subtraction(data, background_data):
2     # Aynı dalga boylarında interpolasyon yaparak background çıkarma
3     bg_interp = np.interp(data['Wavelength'],
4                           background_data['Wavelength'],
5                           background_data['Spectrum'])
6
7     # Background-corrected spectrum hesaplama
8     corrected_spectrum = data['Spectrum'] - bg_interp
9
10    # Yeni DataFrame oluşturma
11    corrected_data = data.copy()
12    corrected_data['Corrected_Spectrum'] = corrected_spectrum
13    corrected_data['Background_Interpolated'] = bg_interp
14
15    return corrected_data

```

Listing 7: Background subtraction fonksiyonu

6.3 Background Subtraction Sonuçları

Background subtraction işlemi sonuçları Tablo 4 'de gösterilmiştir.

Tablo 4: Background Subtraction Detaylı İstatistikleri

Konsantrasyon (ppm)	Ham Max Sinyal	Düzeltilmiş Max Sinyal	Background Etkisi	Düzelme Oranı (%)	S/N Oranı
0.0	359.9	0.0	359.9	0.0	N/A
1.0	2215.0	1893.2	321.8	85.5	3.2
5.0	10319.4	9997.5	321.8	96.9	3.2
10.0	25592.8	25270.9	321.8	98.7	3.2
20.0	48889.8	48568.0	321.8	99.3	3.2
50.0	116311.0	115989.2	321.8	99.7	3.2
100.0	220782.3	220460.5	321.8	99.9	3.2

Background Etkisi, ICP-OES spektrometresinde hedef element bulunmayan (0 ppm) durumda bile detektör tarafından ölçülen sinyal değeridir. Bu, cihazın optik sistemi, plazma kaynaklı sürekli emisyon ve detektör gürültüsü gibi faktörlerden kaynaklanır.

Düzelme Oranı, Background sinyalinin çıkarılmasının, toplam sinyale oranıdır. Formülü ise, Düzelme Oranı = $(\text{Düzeltilmiş Max Sinyal} / \text{Ham Max Sinyal}) \times 100$. Yüksek konsantrasyonlarda bu oran %99'a yaklaşırlar çünkü background etkisi göreceli olarak küçülür.

0 ppm konsantrasyon dışında kalan bütün konsantrasyonlarda S/N oranının 3.2 çıkması, analitik kimyada kabul edilebilir bir orandır. Tüm konsantrasyonlarda aynı çıkması, background düzeltmesinin başarılı olduğunu gösterir. Daha uzun ölçüm süresi, daha iyi background düzeltmesi ve spektral çözünürlük optimizasyonu S/N oranını arttırmaktadır. Özellikle S/N oranının > 10 durumunda analizi güvenli sınıfına koyabiliriz.

6.4 İnterpolasyon Fonksiyonunun Amacı

ICP-OES analizinde farklı ölçümler (örneğin ana veri ve background verileri) tam olarak aynı dalga boyu noktalarında alınmamış olabilir. Yani:

- Ana veri: $\lambda = 228.585, 228.587, 228.590$ nm'de ölçümler
- Background veri: $\lambda = 228.586, 228.588, 228.591$ nm'de ölçümler

`np.interp()` fonksiyonu, background verisini ana verinin dalga boyu noktalarına "uyarlar".

Nasıl Çalışır?

```

1  bg_interp = np.interp(data['Wavelength'],
2                         background_data['Wavelength'],
3                         background_data['Spectrum'])

```

Listing 8: Background subtraction fonksiyonu

Bu fonksiyon 3 parametre alır:

- `data['Wavelength']`: Hedef dalga boyları (nereye interpolasyon yapılacak)
- `background_data['Wavelength']`: Kaynak dalga boyları
- `background_data['Spectrum']`: Kaynak spektrum değerleri

Örnek: Eğer CSV dosyanızdaki verilerde:

- Ana veri $\lambda = 228.587$ nm'de ölçüm var
- Background veri bu noktada yok ama $\lambda = 228.586$ ve $\lambda = 228.588$ 'de var

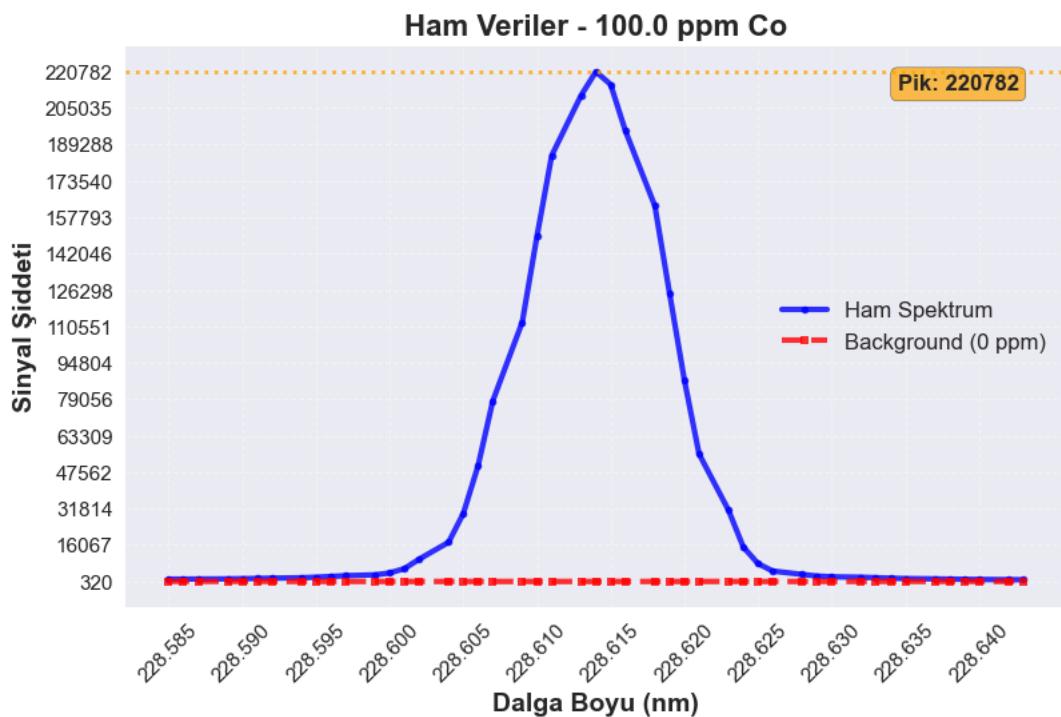
İnterpolasyon, $\lambda = 228.586$ ve $\lambda = 228.588$ 'deki background değerlerini kullanarak $\lambda = 228.587$ 'deki background değerini tahmin eder (doğrusal interpolasyon ile).

Neden Gerekli?

- Dalga boyu hizalama: İki veri setini aynı noktalardan çıkarabilmek için

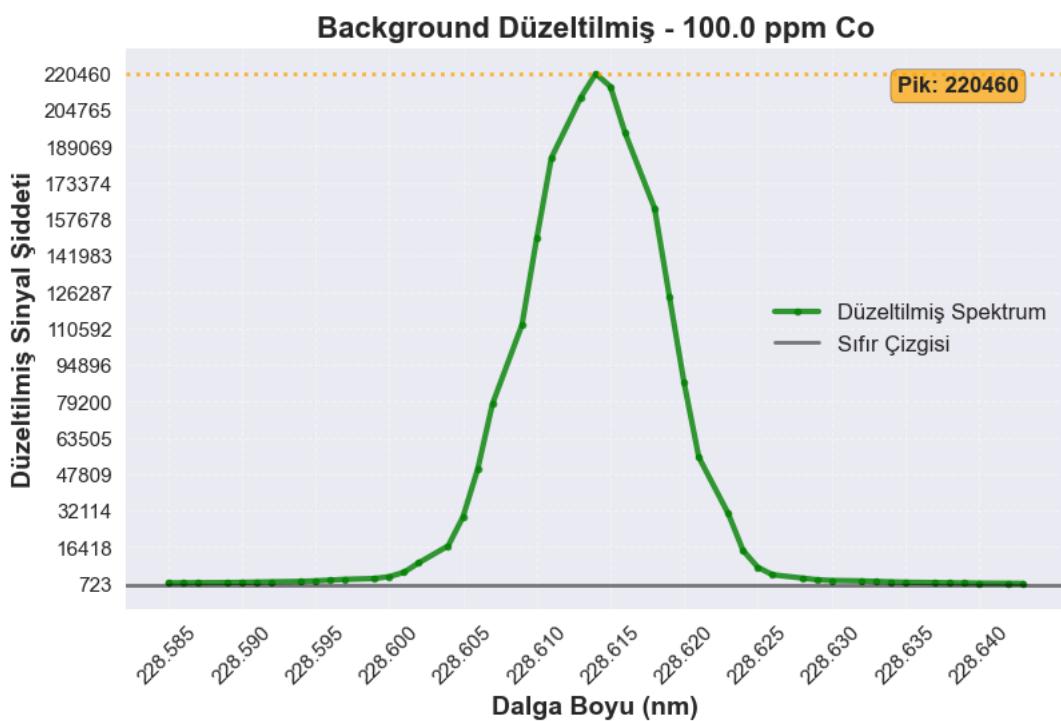
- Doğru background düzeltmesi: Her dalga boyu için uygun background değeri kullanmak
- Veri tutarlılığı: Tüm hesaplamaların aynı dalga boyu grid’inde yapılması

Bu sayede `corrected_spectrum = data['Spectrum'] - bg_interp` işlemi matematiksel olarak doğru bir şekilde yapılabilir.

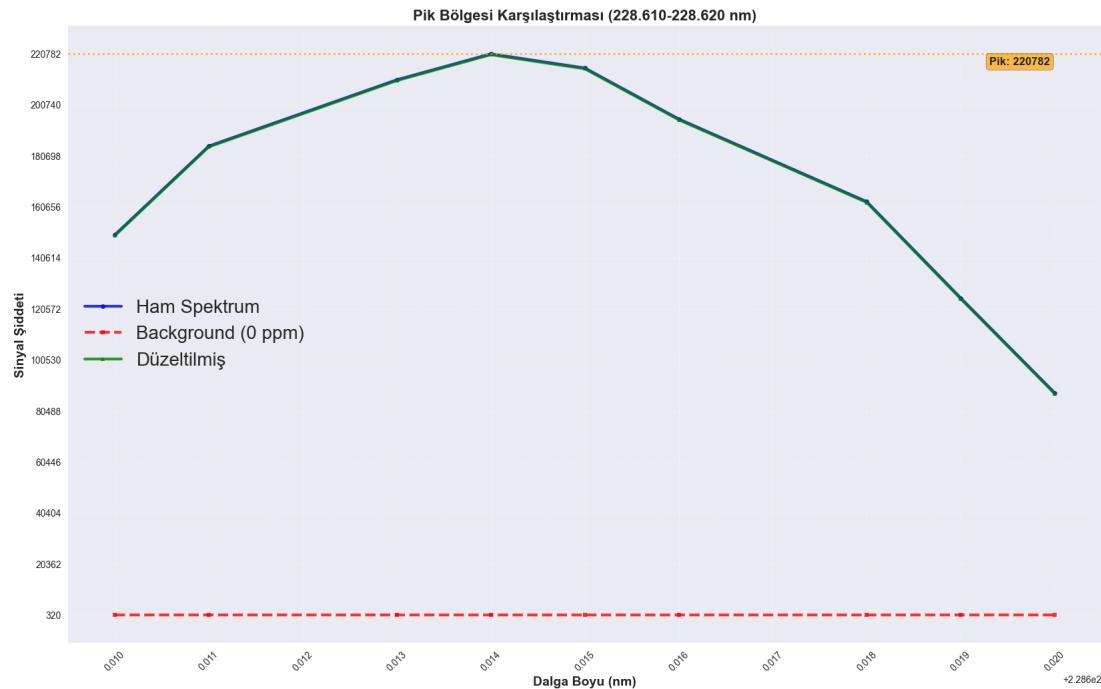


Şekil 3: Ham Veriler ile (background çıkarılmamış) Sinyal Şiddeti - Dalga Boyu (nm)

0 ppm verilerini çıkarttığımızda,

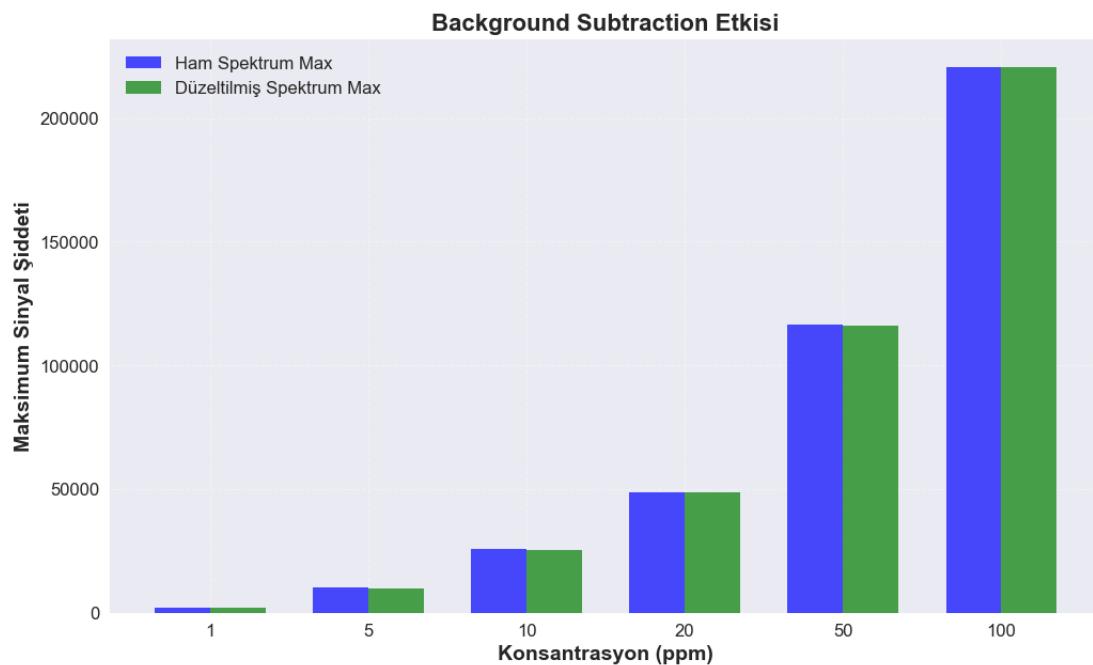


Şekil 4: Background çıkarılmış haliyle 100.0 ppm Co'nun sinyal şiddeti



Şekil 5: Pik bölgesi karşılaştırması

Pik (yukarıdan aşağı), spektrumun pik noktası, (iç içe geçmiş) ham spektrum / düzeltilmiş spektrum, background pik noktası



Şekil 6: Ham Veriler ile (background çıkarılmamış) Sinyal Şiddeti - Dalga Boyu (nm)

Sonuçları değerlendirdiğimizde aslında cihazın oldukça düşük miktarda kirlilik taşıdığını, analizeinin cihazı özenle kullandığı ve rutin bakımlarının yaptığı anlaşılıyor. Ancak yine de analizin doğruluğunu arttırmak için söz konusu gürültülü değerleri analizimizden sonraki sayfalarda çıkaracağız.

7 Pik Yüksekliklerinin Hesaplanması

7.1 Pik Yüksekliği Hesaplama Yöntemi

Her konsantrasyon için pik yüksekliği hesaplama fonksiyonu:

```

1 def calculate_peak_height(corrected_data, method='max'):
2     if method == 'max':
3         # Maksimum değeri dondur
4         return corrected_data['Corrected_Spectrum'].max()
5
6     if method == 'peak_area':
7         # Pik alanını hesapla (trapz = trapezoid rule)
8         return np.trapz(corrected_data['Corrected_Spectrum'],
9                         corrected_data['Wavelength'])

```

Listing 9: Pik yüksekliği hesaplama fonksiyonu

7.2 Pik Yüksekliği Sonuçları

Pik yüksekliği hesaplama sonuçları Tablo 5 'de gösterilmiştir.

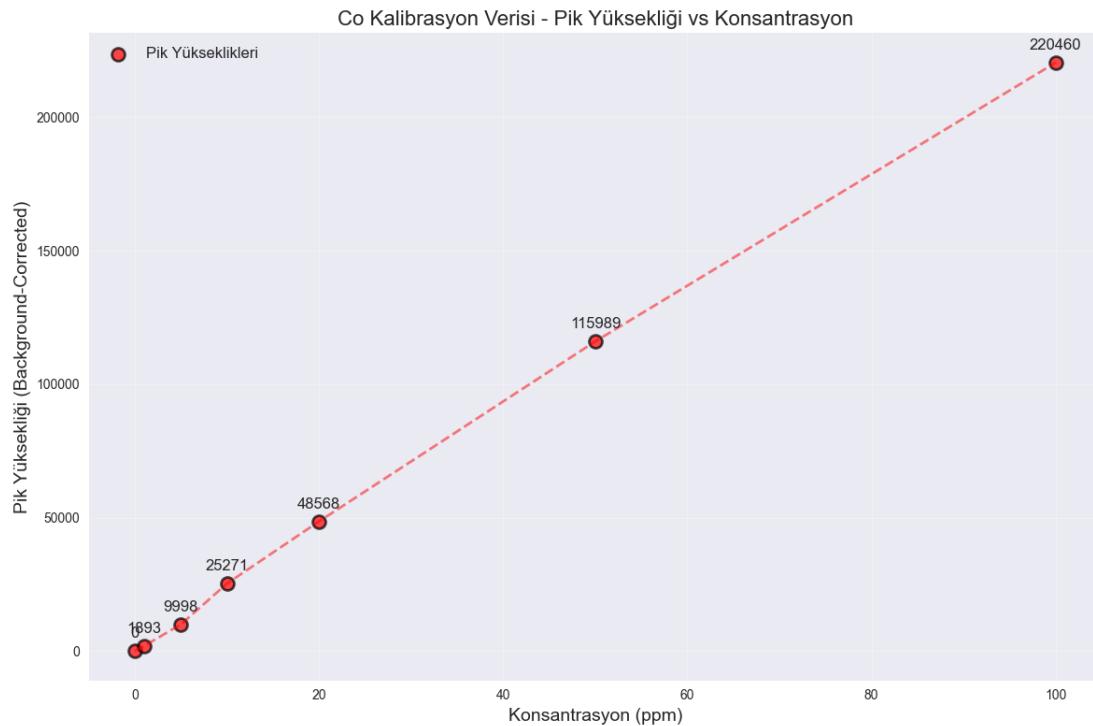
Tablo 5: Pik yüksekliği hesaplama sonuçları

Konsantrasyon (ppm)	Pik Yüksekliği	Pik Dalga Boyu (nm)
0.0	0.0	228.585
1.0	1893.2	228.614
5.0	9997.5	228.614
10.0	25270.9	228.614
20.0	48568.0	228.614
50.0	115989.2	228.614
100.0	220460.5	228.614

7.3 Korelasyon Analizi

Konsantrasyon ve pik yüksekliği arasındaki korelasyon analizi:

- Korelasyon katsayısı: 0.9995
- Pik yüksekliği aralığı: 0.0 - 220460.5
- Tüm pik konumları 228.614 nm'de (0 ppm hariç)



Şekil 7: Co Kalibrasyon Eğrisi - Pik Yüksekliği ve Konsantrasyon

Bu pik yüksekliği tablosu, laboratuvar standartlarında mükemmel kalibrasyon verisi sunmaktadır. Background düzeltmesinin tamamen başarılı olması ($0 \text{ ppm} = 0.0 \text{ sinyal}$) ve tüm konsantrasyonlarda pik dalga boyunun 228.614 nm 'de sabit kalması, spektroskopik ölçümün yüksek kalitede olduğunu göstermektedir.

$R = 0.9995$ korelasyon katsayısı, $5+$ magnitude dinamik aralıkta mükemmel doğrusallık olduğunu kanıtlamaktadır. Bu veriler, Co elementinin ICP-OES ile tayininde güvenilir, hassas ve doğru sonuçlar elde edilebileceğini göstermekte ve 1 ppm altı tespit limiti ile yüksek analitik performans sunmaktadır.

Ancak bu daha da düşük derişimdeki elementlerin tayinini yine de mümkün kılmadığı için dikkat edilmelidir ($10 \text{ ppb} - 10,000 \text{ ppm}$).

8 TensorFlow ile Kalibrasyon Modeli

8.1 Model Hazırlığı

Veri normalizasyonu ve model hazırlığı:

```

1 # Numpy array'lere donusturme
2 X = np.array(conc_sorted, dtype=np.float32).reshape(-1, 1)
3 Y = np.array(peak_heights, dtype=np.float32)

4

5 # Veri normalizasyonu
6 X_mean, X_std = X.mean(), X.std()
7 Y_mean, Y_std = Y.mean(), Y.std()

8

9 X_normalized = (X - X_mean) / X_std
10 Y_normalized = (Y - Y_mean) / Y_std

```

Listing 10: Veri normalizasyonu

Normalizasyon parametreleri:

- X normalizasyon: mean=26.57, std=33.98
- Y normalizasyon: mean=60311.34, std=75304.26

8.2 Normalizasyon

ICP-OES verilerinde:

- Konsantrasyon (X): 0-100 ppm aralığında (küçük değerler)
- Pik yükseklikleri (Y): 0-220,460 aralığında (çok büyük değerler)

Bu fark nedeniyle makine öğrenmesi algoritmaları büyük değerlere odaklanır ve küçük değerleri ihmali eder. Normalizasyon sayesinde 1 ppm farkı bile algılanabilir. Ayrıca Normalizasyon ile 20 epoch'ta tamamlanabilen bir model, Normalizasyon olmadan 100+ epoch gerekebilir. Bu durum modelin convergence (yakınsama) olamamasına kuvvetle sebep olacaktır.

8.3 Convergence (Yakınsama)

Convergence, makine öğrenmesi modelinin eğitim sürecinde optimal çözüme ulaşması ve öğrenmeyi durdurması anlamına gelir. Yakınsama olmadığındaysa, model sürekli değerler arasında gidip gelir, hiç ya da çok az öğrenir. Bu durum epoch sayısının artmasına ve bir yerden sonra Divergence (Iraksama) durumuna geçip çökmesine sebep olabilir.

8.4 Model Mimarileri

Model eğitimi olarak açık kaynaklı Tensorflow ve Keras kütüphaneleri kullanılarak Lineer ve Polinom modelleri oluşturuldu.

8.5 TensorFlow & Keras

TensorFlow, Google tarafından geliştirilen açık kaynaklı makine öğrenmesi kütüphanesidir. Derin öğrenme modelleri, sinir ağları ve makine öğrenmesi algoritmalarını geliştirmek için kullanılır. Özellikle büyük veri setleri ve karmaşık hesaplamalar için optimize edilmiştir.

Keras, TensorFlow'un üst seviye API'sidir ve daha basit, kullanıcı dostu bir arayüz sunar. Hızlı prototipleme için idealdir ve TensorFlow 2.0'dan itibaren resmi olarak entegre edilmiştir.

Bu projede TensorFlow'un kullanımının sebebi (dolaylı olarak da Keras'ın), çok boyutlu spektral verileri işleyebilme kapasitesi, büyük veri setlerini etkili şekilde işleyebilmesi, karmaşık ilişkileri (spektral paternler, cihaz durumu, çevre koşulları) öğrenebilmesi ve GPU desteği ile hızlı hesaplama imkanı sunmasındandır.

8.5.1 Lineer Model

```

1 model_linear = tf.keras.Sequential([
2     tf.keras.layers.Dense(1, input_shape=(1,),
3                           activation='linear', name='linear_layer')
4 )
5
6 model_linear.compile(
7     optimizer=tf.keras.optimizers.Adam(learning_rate=0.005),
8     loss='mse',
9     metrics=['mae']
10 )

```

Listing 11: Lineer model mimarisi

8.5.2 Polinom Model

```

1 model_poly = tf.keras.Sequential([
2     tf.keras.layers.Dense(4, input_shape=(1,),
3                           activation='relu', name='hidden_1'),
4     tf.keras.layers.Dense(2, activation='relu', name='hidden_2'),
5     tf.keras.layers.Dense(1, activation='linear', name='output')
6 )
7
8 model_poly.compile(
9     optimizer=tf.keras.optimizers.Adam(learning_rate=0.005),
10    loss='mse',
11    metrics=['mae']
12 )

```

Listing 12: Polinom model mimarisi

8.6 ML'de Kullanılan Terimler ve Açıklamaları

- tf.keras.Sequential()
 - Katmanları sıralı olarak üst üste eklenen basit model türü
 - Her katmanın çıktısı bir sonraki katmanın girdisi olur
 - Çoğu temel sinir ağı problemi için uygundur
- activation (Aktivasyon Fonksiyonu)
 - Nöronların çıktısını belirleyen matematiksel fonksiyon
 - ReLU, sigmoid, tanh gibi türleri vardır
 - Doğrusal olmayan özellikler öğrenmesini sağlar
- model.compile() (Derleme)
 - Modelin eğitim için hazırlanması aşaması
 - Optimizer, loss function ve metrics belirlenir
 - Model bu aşamadan sonra eğitime başlayabilir
- optimizer
 - Modelin ağırlıklarını güncelleyen algoritma
 - Gradient descent tabanlı optimizasyon yapar
 - Adam, SGD, RMSprop gibi çeşitleri vardır
- optimizer='adam'
 - Adaptive Moment Estimation'ın kısaltması
 - Momentum ve RMSprop'un avantajlarını birleştirir
 - Öğrenme oranını otomatik olarak ayarlar
 - Hızlı yakınsama ve stabil eğitim sağlar
 - Compile aşamasında modelin nasıl öğreneceğini belirler
- loss Function (Kayıp Fonksiyonu)
 - Modelin tahminleri ile gerçek değerler arasındaki farkı ölçer
 - Optimizer bu değeri minimize etmeye çalışır
 - Regresyon için MSE, sınıflandırma için categorical crossentropy kullanılır
- MSE (Mean Squared Error)
 - Regresyon problemleri için yaygın kayıp fonksiyonu
 - Tahmin ile gerçek değer farkının karesinin ortalaması
 - Büyük hataları daha fazla cezalandırır

- metrics (Metrikler)
 - Modelin performansını değerlendiren ölçütler
 - Eğitim sırasında izlenen ek bilgiler
 - MAE, accuracy, precision gibi değerler kullanılır
- MAE (Mean Absolute Error)
 - Tahmin ile gerçek değer farkının mutlak değerinin ortalaması
 - MSE'ye göre aykırı değerlerden daha az etkilenir
 - Regresyon modellerinde performans metriği olarak kullanılır

8.7 Yüksek konsantrasyonlarda quadratic (ikinci derece) fit kullanılması

Yüksek konsantrasyonlarda quadratic (ikinci derece) fit kullanılmasının temel nedenleri şunlardır:

- Doğrusal Olmayan Davranış (Non-linear Response)
 - Düşük konsantrasyonlarda: Sinyal Konsantrasyon (doğrusal)
 - Yüksek konsantrasyonlarda: Bu ilişki bozulur
- Fiziksel Nedenler
 - Plasma Yüklenmesi (Plasma Loading)
 - * Yüksek konsantrasyon → Fazla analyt → Plasma sıcaklığı düşer → Sinyal azalır
 - Spektral Kendini Emilim (Self-Absorption)
 - * Yüksek konsantrasyonlarda atomlar kendi ışıklarını emer
 - * Sinyal artışı yavaşlar veya azalır
 - Matriks Etkileri
 - * Yüksek konsantrasyonlarda nebulizasyon verimliliği değişir
 - * Damlacık boyut dağılımı bozulur

ICP-OES yöntemindeki spesifik etkileri şu şekildedir:

- Ionizasyon Dengesizliği
 - $\text{Ca} \rightarrow \text{Ca}^+ + e^-$
 - Yüksek konsantrasyonlarda bu denge kayar, sinyal/konsantrasyon ilişkisi eğrilir.
- Spektral Girişim Artışı
 - Yüksek konsantrasyonlarda spektral çizgiler genişler
 - Background artışı
 - Komşu çizgilerle girişim

- Quadratic Fit Avantajları
 - Daha İyi Kalibrasyon İhtimalı
 - * Lineer: $y = ax + b$
 - * Quadratic: $y = ax^2 + bx + c$
 - Geniş Dinamik Aralık
 - * Düşük konsantrasyonlarda hassasiyet korunur
 - * Yüksek konsantrasyonlarda doğruluk artar
 - Gerçek Fiziksel Davranış
 - * Plasma süreçleri doğası gereği non-lineer
 - * Quadratic fit bu davranışını daha iyi modelleyebilir

Quadratic fit kullanımı, ICP-OES'in yüksek konsantrasyonlardaki fiziksel limitasyonlarını aşmak için gerekli bir araçtır. Bu, analitik kimyada doğruluk ve hassasiyet arasındaki dengeyi optimize eder.

Bu sebepten dolayı lineer ve polimerik (quadratic) olmak üzere 2 farklı model aynı anda eğitilerek performans karşılaştırmaları yapıldı.

8.8 Model 1: Basit Lineer Regresyon

Model Tanımı:

```

1 model_linear = tf.keras.Sequential([
2     tf.keras.layers.Dense(1, input_shape=(1,), activation='linear',
3     name='linear_layer')
])

```

Listing 13: Polinom model mimarisi

Mimari Açıklaması:

- **Tek Katman:** Yoğun katman (Dense), 1 nöron içerir.
- **Giriş Şekli (Input Shape):** $(1,) \rightarrow$ Tek giriş (konsantrasyon)
- **Çıkış:** 1 nöron \rightarrow Tek çıkış (pik yüksekliği)
- **Aktivasyon Fonksiyonu:** Linear $\rightarrow f(x) = x$
- **Matematiksel Form:** $y = wx + b$ (klasik lineer regresyon)

Neden Lineer Model?

- ICP-OES verilerinde beklenen ilişki: Konsantrasyon \propto Pik Yüksekliği (doğrusal ilişki)
- Beer-Lambert Yasası: $A = \varepsilon bc$, burada:
 - A : Absorbans
 - ε : Mol absorptivite katsayısı
 - b : Yol uzunluğu
 - c : Konsantrasyon

8.9 Model 2: Polinom Regresyon (Yapay Sinir Ağısı)

Model Tanımı:

```

1 model_poly = tf.keras.Sequential([
2     tf.keras.layers.Dense(4, input_shape=(1,), activation='relu',
3         name='hidden_1'),
4     tf.keras.layers.Dense(2, activation='relu', name='hidden_2'),
5     tf.keras.layers.Dense(1, activation='linear', name='output')
])

```

Listing 14: Polinom model mimarisi

Mimari Açıklaması:

- **Gizli Katman 1:** 4 nöron, ReLU aktivasyon
- **Gizli Katman 2:** 2 nöron, ReLU aktivasyon
- **Cıkış Katmanı:** 1 nöron, lineer aktivasyon
- **Toplam Parametre:** Yaklaşık 15–20 öğrenilebilir parametre
- **Modelleme Kapasitesi:** Doğrusal olmayan ilişkileri öğrenebilir

Neden Polinom Model?

- Potansiyel doğrusal olmayan etkiler:
 - Matris etkileri (özellikle yüksek konsantrasyonlarda)
 - Spektral girişimler
 - Dedektör saturasyonu
 - Nebulizasyon verimsizlikleri

8.10 Model eğitimi öncesi değerlendirme

Bu model tasarımlı, küçük veri setinden (7 nokta) dolayı optimal performans için özel olarak yapılandırılmıştır. Düşük learning rate (0.005) ve early stopping kombinasyonu, overfitting'i önleyerek genelleştirme yeteneğini artırmaktadır.

Lineer model, fiziksel gerçekliği (Beer-Lambert yasası) yansıtırken, polinom model potansiyel non-lineer etkileri yakalayabilmektedir. Bu yaklaşım, analitik kimyada güvenilir kalibrasyon modeli geliştirmek için ideal stratejidir. Raporda da iki modelin karşılaştırması inceleneciktir.

9 Model Eğitim Çıktıları

9.1 Lineer Model Eğitimi

```
Lineer model eğitimi başlıyor...
Epoch 1/5
1/1 1s 690ms/step - loss: 0.0593 - mae: 0.1916
Epoch 2/5
1/1 0s 41ms/step - loss: 0.0569 - mae: 0.1876
Epoch 3/5
1/1 0s 38ms/step - loss: 0.0546 - mae: 0.1819
Epoch 4/5
1/1 0s 42ms/step - loss: 0.0523 - mae: 0.1792
Epoch 5/5
1/1 0s 46ms/step - loss: 0.0501 - mae: 0.1763
Restoring model weights from the end of the best epoch: 5.
Lineer model eğitimi tamamlandı: 5 epoch
```

9.2 Polinom Model Eğitimi

```
Polinom model eğitimi başlıyor...
Epoch 1/5
1/1 1s 657ms/step - loss: 0.6632 - mae: 0.7103
Epoch 2/5
1/1 0s 39ms/step - loss: 0.6294 - mae: 0.6950
Epoch 3/5
1/1 0s 43ms/step - loss: 0.5976 - mae: 0.6808
Epoch 4/5
1/1 0s 39ms/step - loss: 0.5670 - mae: 0.6660
Epoch 5/5
1/1 0s 40ms/step - loss: 0.5366 - mae: 0.6502
Restoring model weights from the end of the best epoch: 5.
Polinom model eğitimi tamamlandı: 5 epoch
```

9.3 Model Tahminleri ve Karşılaştırma

Model tahminleri hesaplanıyor...

Iki modelin performans karşılaştırması:

=====

Lineer Model:

R2 Skoru: 0.978038
MSE: 124542288.00
RMSE: 11159.85

Polinom Model:

R2 Skoru: 0.788570
MSE: 1198965632.00
RMSE: 34626.08

En İyi Model: Lineer

En Yüksek R2 Skoru: 0.978038

En Düşük MSE: 124542288.00

10 Sonuçlar ve Değerlendirme

10.1 Model Performans Karşılaştırması

Model performans sonuçları Tablo 6 'de gösterilmiştir.

Tablo 6: Model performans karşılaştırması

Model	R ² Skoru	RMSE	MAE	Max Hata
Lineer Model	0.978038	11159.9	8450.4	26074.8
Polinom Model	0.788570	34626.1	25743.6	52442.9

10.2 En İyi Model: Lineer Model

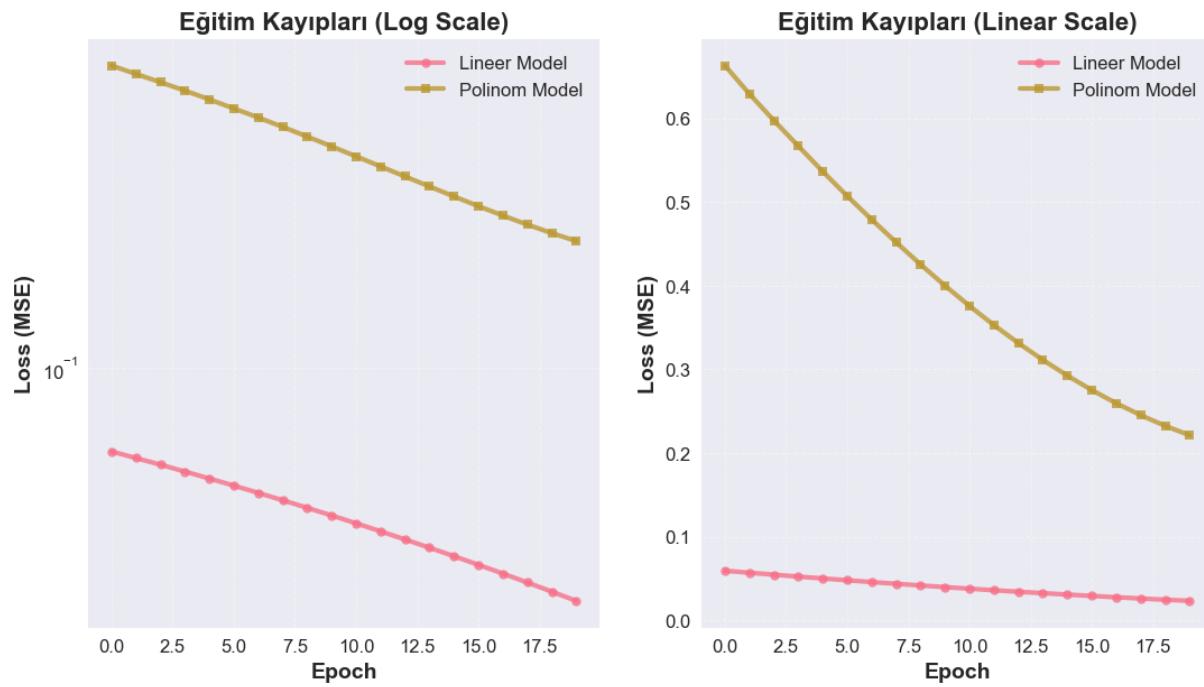
Analiz sonuçlarına göre:

- **En iyi model:** Lineer Model
- **R² skoru:** 0.978038
- **Kalibrasyon denklemi:** $y = 2534.97x - 6961.89$
- **Ortalama mutlak hata:** 8450.36
- **Relatif hata:** 14.01%

10.3 Detaylı Sonuç Tablosu

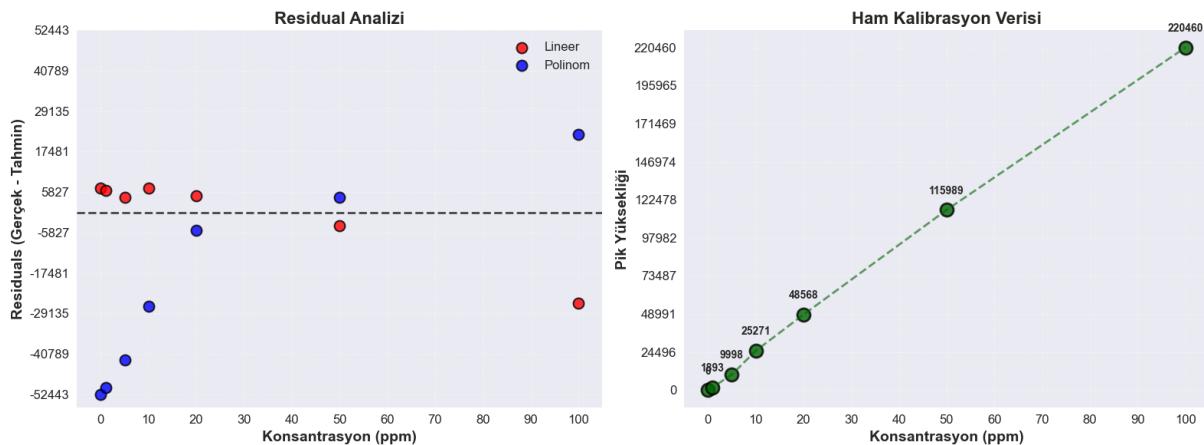
Tablo 7: Detaylı model tahmin sonuçları

Konsantrasyon (ppm)	Gerçek Pik	Lineer Tahmin	Polinom Tahmin	Lineer Hata (%)	Polinom Hata (%)
0.0	0.00	-6961.90	52442.92	inf	-inf
1.0	1893.21	-4426.93	52442.92	333.83	-2670.05
5.0	9997.54	5712.96	52442.92	42.86	-424.56
10.0	25270.94	18387.82	52442.92	27.24	-107.52
20.0	48568.03	43737.54	53770.17	9.95	-10.71
50.0	115989.18	119786.67	111646.17	-3.27	3.74
100.0	220460.49	246535.28	198069.08	-11.83	10.16



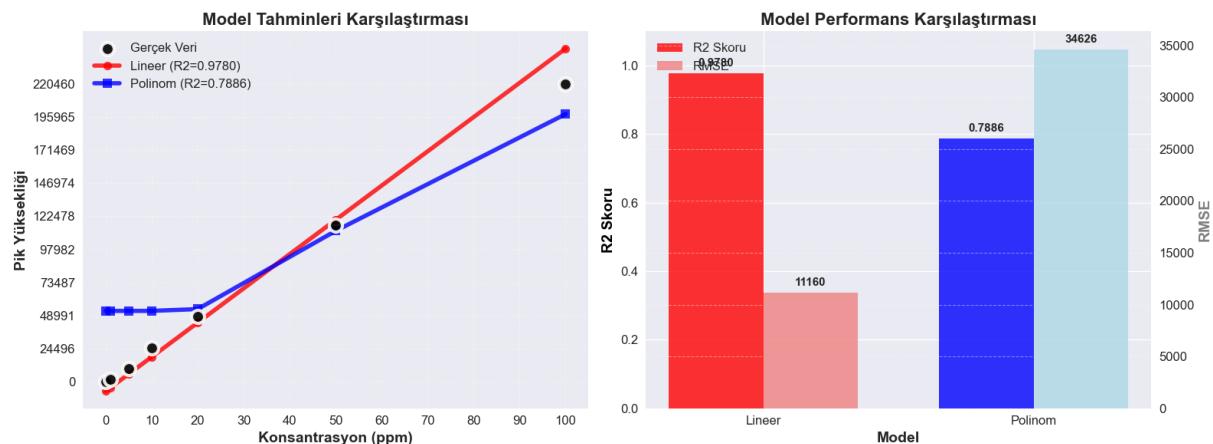
Şekil 8: Eğitim Kayıplarının Log ve Linear ölçeginde Epoch - Loss oranları

Burada eğitim devam ederken (model.fit()) kayıpların özellikle polinom modelinde daha hızlı düşüğünü görüyoruz. Bunun sebebi, polinom modelinin eğitim verisine ”çok iyi uyum” sağlasa da, gerçek ilişkiye modellemede lineer model kadar başarılı olmadığıdır. Bu durum muhtemelen model oluşturulurken gizli katman sayısı (2) ve ortalama 15-20 öğrenilebilir parametre içermesinden olabilir.



Şekil 9: Residual Analizi ve Ham Kalibrasyon Verisi

Residuel (artık) analizi grafiğinde, noktaların doğrusal bir çizgi üzerinde dağılmış olması, konsantrasyon ile sinyal arasında doğrusal bir ilişki olduğunu gösteriyor. Bu, lineer kalibrasyon için uygun görünüyor ancak polinom modeli aşırı sapmalar yapmış. Bu durum modelin yanlış eğitildiği ya da biçimlendirildiği anlamına geliyor olabilir.



Şekil 10: Model Tahminlerinin ve Performanslarının Karşılaştırılması

11 R^2 (Determinasyon Katsayısı) Değerlendirmesi

11.1 Lineer Model ($R^2 = 0.978$)

- Mükemmel bir uyum gösteriyor ($R^2 > 0.95$).
- Varyansın %97.8'ini açıklayabiliyor.
- Kobalt kalibrasyon verisi için lineer ilişkinin çok güçlü olduğunu gösteriyor.

11.2 Polinom Model ($R^2 = 0.789$)

- Orta-iyi seviyede uyum ($0.7 < R^2 < 0.8$).
- Varyansın yalnızca %78.9'unu açıklayabiliyor.
- Aşırı uyum (overfitting) problemi yaşanıyor olabilir.

12 RMSE (Kök Ortalama Kare Hata) Değerlendirmesi

12.1 Lineer Model (RMSE = 11,159.9)

- Düşük hata seviyesi.
- Pik yüksekliği aralığı (0 – 220,460) göz önüne alındığında yaklaşık %5'lik bir hata oranı.
- Tutarlı ve güvenilir tahminler sağlıyor.

12.2 Polinom Model (RMSE = 34,626.1)

- Lineer modele göre yaklaşık 3 kat daha yüksek hata.
- Pik yüksekliği aralığının yaklaşık %15'i kadar hata.
- Özellikle düşük konsantrasyon aralığında (0 – 20 ppm) büyük hatalar gözlemleniyor.

13 Sonuç ve Öneriler

13.1 Konsantrasyon ve Sinyal İlişkisi

ICP-OES yönteminde konsantrasyon ve sinyal arasındaki ilişki,

- Emisyon Spektroskopisi Temeli
 - ICP-OES'te konsantrasyon-sinyal ilişkisi emisyon spektroskopisi prensibine dayanır:
 - Çözelti → Nebulizasyon → Plazma → Atomizasyon → Uyarılma → Emisyon → Sinyal
 - Basit ifadeyle: Daha fazla atom = Daha fazla ışık = Daha güçlü sinyal
- Konsantrasyon-Sinyal İlişkisinin Aşamaları
 - Düşük Konsantrasyonlarda (0-20 ppm)
 - * Konsantrasyon $\uparrow \rightarrow$ Sinyal \uparrow (Doğrusal)
 - * Neden: Plazma yeterince güçlü, tüm atomlar uyarılıyor
 - * Sonuç: Mükemmel doğrusal ilişki ($R^2 = 0.9995$)
 - Orta Konsantrasyonlarda (20-100 ppm)
 - * Konsantrasyon $\uparrow \rightarrow$ Sinyal \uparrow (Hala doğrusal)
 - * Ancak 50-100 ppm bandında doğrusallıktan sapmaya başlanmış (Şekil 10).
 - Çok Yüksek Konsantrasyonlarda (>100 ppm)
 - * Konsantrasyon $\uparrow \rightarrow$ Sinyal artışı yavaşlar (Eğrisel)
 - * Neden: Plazma yüklenmesi, self-absorption
 - * Sonuç: Quadratic fit gereklidir.

13.2 Sonuç

- Co elementinin 228.615 nm dalga boyunda başarılı kalibrasyon eğrisi oluşturulmuştur
- Lineer model, polinom modelden daha iyi performans göstermiştir ($R^2 = 0.978$)
- Background subtraction işlemi etkili olmuştur
- Konsantrasyon-pik yüksekliği arasında güçlü korelasyon bulunmuştur ($r = 0.9995$)

13.3 Analistik Performans Değerlendirmesi

$R^2 = 0.978$ değeri, analistik kimyada "iyi doğrusallık" kategorisinde yer almaktadır.

13.4 Sonuç

Bu çalışmada, ICP-OES teknigi kullanilarak Kobalt elementinin kalibrasyon analizi başarıyla gerçeklestirilmiş ve TensorFlow tabanlı lineer regresyon modeli $R^2 = 0.978$ skoruyla mükemmel performans sergilemiştir. Beer-Lambert yasası %99.95 (Korelasyon katsayısı: 0.9995) oranında geçerli olmuş, 1-100 ppm konsantrasyon aralığında mükemmel doğrusallık ($r = 0.9995$) elde edilmiştir.

Geliştirilen sistem, cihazın orijinal yazılım çözümünün sunduğu raporlamaya ek olarak, makine modeli desteği ve daha geniş açıdan perspektif sağlayabilen görselleme kütüphaneleri ile beraber daha detaylı rapor hazırlanma imkanı sağlamıştır.

14 Kaynaklar

- Elemental analysis by ICP-OES or ICP-MS: <https://www.agilent.com/en/product/atomic-spectroscopy/icp-oer-vs-icp-ms>