**1. Key Differences Between the Two Code Snippets**

**a. Data Processing Pipeline**

* **First Code Snippet**:
  + Uses process\_all\_files, which introduces noise, adds anomalies, preprocesses the data, balances it, and simulates predictions.
  + The evaluation is based on simulated predictions (Label column) compared to the ground truth (True\_Label).
  + The accuracy is calculated using evaluate\_classification, which compares True\_Label and Label.
* **Second Code Snippet**:
  + Uses process\_with\_dmd, which applies Dynamic Mode Decomposition (DMD) for anomaly detection.
  + The accuracy is calculated by comparing the DMD-based anomaly labels (DMD\_Label) with the ground truth (True\_Label).

**b. Anomaly Detection Mechanism**

* **First Code Snippet**:
  + Anomalies are introduced explicitly, and the evaluation is based on simulated predictions with a mislabeling probability.
  + The mislabeling probability is inversely proportional to SNR, which might lead to fewer mislabels at higher SNR values, resulting in higher accuracy.
* **Second Code Snippet**:
  + Anomalies are detected using DMD, which reconstructs the data and identifies anomalies based on reconstruction error.
  + The DMD-based approach might not be as effective in detecting the specific types of anomalies introduced (e.g., drops and spikes), leading to lower accuracy.

**2. Potential Reasons for the Discrepancy**

**a. DMD's Sensitivity to Anomalies**

* DMD is a linear dimensionality reduction technique that assumes the data can be approximated by a low-rank linear model. If the anomalies introduced (e.g., drops and spikes) do not significantly affect the low-rank structure of the data, DMD might fail to detect them effectively.
* The reconstruction error threshold (95th percentile) might not be optimal for detecting the specific anomalies in your dataset.

**b. Mislabeling Simulation vs. DMD Detection**

* In the first snippet, the mislabeling simulation is controlled and directly tied to the SNR. At high SNR values, the mislabeling probability is low, leading to high accuracy.
* In the second snippet, DMD's ability to detect anomalies depends on how well the anomalies deviate from the normal data structure. If the anomalies are subtle or do not significantly impact the reconstruction error, DMD might miss them.

**c. Data Balancing**

* The first snippet uses SMOTE to balance the dataset, which might improve the performance of the evaluation metrics.
* The second snippet does not explicitly balance the dataset before applying DMD, which could lead to biased results if the dataset is imbalanced.

**d. SNR Values**

* The SNR values in the two snippets are generated differently:
  + First snippet: snr\_values = sorted([random.uniform(10, 1000) for \_ in range(10)])
  + Second snippet: snr\_values\_dmd = sorted([random.uniform(10, 1000) for \_ in range(15)], reverse=True)
  + The range and distribution of SNR values might affect the results, especially if DMD performs differently at different SNR levels.

**3. Conclusion**

The discrepancy in accuracy is likely due to differences in how anomalies are detected and evaluated in the two approaches. DMD might not be as effective in detecting the specific types of anomalies in your dataset, especially if they do not significantly impact the low-rank structure of the data. By analyzing the DMD results, tuning its parameters, and comparing the reconstruction error with the ground truth, you can identify the root cause and improve the accuracy of the DMD-based approach. Let me know if you need further assistance!