Contents

[Case western data (appending normal + faulty) in SINDy 2](#_Toc111792299)

[NASA bearing datasets 5](#_Toc111792300)

[Pre-processing the data 5](#_Toc111792301)

[NASA bearing data in SINDy 10](#_Toc111792302)

# Case western data (appending normal + faulty) in SINDy

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | Appending case western faulty data to normal operation data | | |
| **Test** | **Output** | | **Comment** |
| SINDy  w/ Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5'] | SINDy Model  (x0)' = 0.005 x0 + -0.061 x0^2 + -0.292 x0^3 + 0.386 x0^4 + 3.731 x0^5 | | (in normal\_sindy.py)  No noticeable change at any point – so cant see when the fault starts (would expect it halfway) |
|  | SINDy Model  (x0)' = 0.004 x0 + -0.064 x0^2 + -0.132 x0^3 + 0.518 x0^4 | SINDy Model  (x0)' = -0.001 x0 + 0.008 x0^2 + -0.002 x0^3 + -0.012 x0^4 + 0.006 x0^5 | Compare to the normal operation data and the faulty operation data on their own |
| SINDy-PI  Library:  ['1', 'sin(t)', 'x0\_t', 'x0\_tt', '1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']  Threshold = 1e-3 | x0\_t = 0.0003 1 + 0.0013 x0\_tt + 0.0003 1 + 0.0039 x0 + -0.1354 x0^2 + -0.1214 x0^3 + 2.0382 x0^4 | | (in normal\_sindypi.py) |
| Ensemble-SINDy  Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5'] | Ensemble with replacement (V1)  (x0)' = 0.005 x0 + -0.061 x0^2 + -0.292 x0^3 + 0.386 x0^4 + 3.731 x0^5    Ensemble model without replacement (V2)  (x0)' = 0.005 x0 + -0.061 x0^2 + -0.292 x0^3 + 0.386 x0^4 + 3.731 x0^5 | | V1 and V2 are the same – unsure whether need to fix that (or if they just really are the same?)  Tried re making the model before fitting with V2 ensembling but this makes the equation = 0.000 |
| Ensemble-SINDy  Ensembling version V1-4  Threshold = 0.00 | Ensemble with replacement (V1)  (x0)' = 0.001 1 + 0.004 x0 + -0.141 x0^2 + -0.136 x0^3 + 2.253 x0^4 + 0.127 x0^5    Ensemble model without replacement (V2)  (x0)' = 0.001 1 + 0.004 x0 + -0.141 x0^2 + -0.136 x0^3 + 2.253 x0^4 + 0.127 x0^5    Ensemble library (V3)  (x0)' = 0.001 1 + 0.004 x0 + -0.141 x0^2 + -0.136 x0^3 + 2.253 x0^4 + 0.127 x0^5    Both types of ensembling w/ candidate drops (V4)  (x0)' = 0.001 1 + 0.003 x0 + -0.144 x0^2 + 2.323 x0^4 + -2.503 x0^5 | |  |
| Ensemble-SINDy  Ensembling version V1-4  Threshold = 0.01 | Ensemble with replacement (V1)  (x0)' = -0.055 x0^2 + 0.034 x0^3 + 0.171 x0^4  Ensemble model without replacement (V2)  (x0)' = -0.052 x0^2 + 0.036 x0^3  Ensemble library (V3)  (x0)' = -0.052 x0^2 + 0.036 x0^3  Both types of ensembling w/ candidate drops (V4)  (x0)' = -0.050 x0^2 | | Different models for V1 and V2, but V2 and V3 still the same  V4 still more sparse than 1-3  Only put graph for V1 – they all look the same  Note: changing initial condition from 0 to the first datapoint in x\_train didn’t change anything |

# NASA bearing datasets

## Pre-processing the data

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | Plot NASA bearing datasets: test 1 data | | |
| **Test** | **Output** | (description) | **Comment** |
| Running example from Kaggle: just plotting the data from **1st test** |  | Raw data: amplitude of vibration over time | Refer to Kaggle page to see what each of these plots are meant to be  🡪 Bearing 1 – Ch 1&2; Bearing 2 – Ch 3&4; **Bearing 3 – Ch 5&6**; **Bearing 4 – Ch 7&8**.  🡪each bearing has 2 accelerometers – one for x-axis, one for y-axis  🡪 Description: At the end of the test-to-  failure experiment, inner race defect occurred in **bearing 3** and roller element defect in **bearing 4**.  This took quite long to run  Source for definition of Shannon entropy:  <https://www.sciencedirect.com/>  science/article/pii/B9780128201251000142  Source for definition of kurtosis:  https://www.scribbr.com/statistics/kurtosis/ |
|  | Rms: root mean square of amplitude over time |
|  | Kurtosis (?):  Kurtosis is **a measure of the tailedness of a distribution**. Tailedness is how often outliers occur. |
|  | Shape: rms/ mean value |
|  | Impulse: max value/ mean value |
|  | Crest: max value / rms value |
|  | Entropy:  Shannon entropy is **the average rate at which information is produced by a stochastic source of data** |
|  |  | Plot of the training data |  |
|  |  |  |  |
|  |  |  |  |
| **Task** | Plot NASA bearing datasets: test 2 data | | |
| Plotting the raw data from test 2 |  |  | **Set No. 2:** Recording Duration: February 12, 2004 10:32:39 to February 19, 2004 06:22:39 No. of Files: 984 No. of Channels: 4 Channel Arrangement: Bearing 1 – Ch 1; Bearing2 – Ch 2; Bearing3 – Ch3; Bearing 4 – Ch 4. File Recording Interval: Every 10 minutes File Format: ASCII Description: At the end of the test-to-failure experiment, outer race failure occurred in bearing 1. |

## NASA bearing data in SINDy

|  |  |  |
| --- | --- | --- |
| **Test** | **Output** | **Comment** |
| **Normal operation data** | | |
| Applying test 1 channel 1 (bearing 1, x-axis), no fault occurs on this bearing | Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']  SINDy Model  (x0)' = 0.000 | There’s no fault recorded for this data  Underfit model – probably means the sparsification parameter was set too high |
| **SINDy** |
| **SINDy**  Threshold changed from 5e-4 to 1e-3 | Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']  SINDy Model  (x0)' = 0.249 x0 + -7.316 x0^2 + 67.718 x0^3 + -229.386 x0^4 + 250.693 x0^5 | the model had really high coefficients but the plot looks pretty flat? |
| **SINDy-PI** | Library:  ['1', 'sin(t)', 'x0\_t', 'x0\_tt', '1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']  Model 10  1 = 1.0000 1  sin(t) = 0.0108 1 + 34.2996 x0\_t + 3.9538 x0\_tt + 0.0108 1 + 0.4242 x0 + -4.6838 x0^2  x0\_t = 0.0000  x0\_tt = 0.0018 1 + 0.0018 1 + -0.0279 x0  1 = 1.0000 1  x0 = 0.0305 1 + 0.0305 1 + 4.0870 x0^2  x0^2 = -0.0066 1 + -0.0066 1 + 0.2310 x0  x0^3 = -0.0012 1 + -0.0012 1 + 0.0356 x0  x0^4 = 0.0013 x0  x0^5 = 0.0000  x0^6 = 0.0000 | Not sue what that spike is at the start – try plotting as a scatter diagram rather than a line diagram  -scatter function didn’t work ? |
| **Ensemble- SINDy** | \*\*\*how to ensemble has changed on the github – check\*\* |  |
| **Faulty operation data** | | |
| Applying test 1 channel 5 (bearing 3, x-axis), which has an inner race defect | Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']  SINDy Model  (x0)' = 1.432 1 + -47.708 x0 + 637.320 x0^2 + -4349.359 x0^3 + 15903.593 x0^4 + -29374.067 x0^5 + 21390.425 x0^6 | Note that here ‘time’ isn’t continuous-may have to rethink t for this data |
| **SINDy** |
| **SINDy-PI** | Library:  ['1', 'sin(t)', 'x0\_t', 'x0\_tt', '1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']    Model 10  1 = 1.0000 1  sin(t) = -0.0478 1 + 7.3784 x0\_t + 0.3957 x0\_tt + -0.0477 1 + 1.1584 x0 + -24.4255 x0^3  x0\_t = 0.0025 1 + 0.0001 sin(t) + 0.1869 x0\_tt + 0.0025 1 + -0.0702 x0 + 0.0773 x0^2 + 1.2988 x0^3  x0\_tt = 0.0050 1 + 1.4731 x0\_t + 0.0050 1 + -0.0766 x0 + 0.3374 x0^4  1 = 1.0000 1  x0 = 0.0331 1 + -0.0411 x0\_t + 0.0189 x0\_tt + 0.0331 1 + 4.5657 x0^2 + -6.0468 x0^3  x0^2 = -0.0071 1 + 0.0167 x0\_t + -0.0071 1 + 0.2184 x0 + 1.2754 x0^3  x0^3 = 0.0010 1 + 0.0462 x0\_t + 0.0112 x0\_tt + 0.0010 1 + -0.0615 x0 + 0.4783 x0^2  x0^4 = 0.0443 x0\_t + 0.0116 x0\_tt + -0.0167 x0 + 0.1422 x0^2  x0^5 = -0.0003 1 + 0.0213 x0\_t + 0.0087 x0\_tt + -0.0003 1 + 0.0315 x0^2  x0^6 = -0.0002 1 + 0.0035 x0\_t + 0.0061 x0\_tt + -0.0002 1 + 0.0025 x0 + 0.0039 x0^2 |  |
| **Ensemble-SINDy** | Library:  [1, 'x0', 'x0^2', 'x0^3']    Ensemble with replacement (V1)  (x0)' = 0.210 1 + -5.091 x0 + 40.870 x0^2 + -108.928 x0^3    Ensemble model without replacement (V2)  (x0)' = 0.210 1 + -5.091 x0 + 40.870 x0^2 + -108.928 x0^3 | For ensemble-had to cut x\_train down to 1000 (initially has an error w/out doing this)  -but then I’ve lost the fault occurring at the end |
| Test 2, ch1 bearing 1, outer race failure | Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']    SINDy Model  (x0)' = 0.249 x0 + -7.316 x0^2 + 67.718 x0^3 + -229.386 x0^4 + 250.693 x0^5 | Model doesn’t look as though it as accounted for the spike at the end   * Goes up and comes back down (again might be easier to see if it was scatter not line plot) * Looking at scatter, would consider last point an outlier – try get rid of it? |
| SINDy |
| SINDy-PI | Library:  ['1', 'sin(t)', 'x0\_t', 'x0\_tt', '1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']  Model 0  Model 1  Model 2  Model 3  Model 4  Model 5  Model 6  Model 7  Model 8  Model 9  Model 10  1 = 0.9999 1  sin(t) = 0.0008 1 + -2.9655 x0\_t + 0.5526 x0\_tt + 0.0008 1  x0\_t = -0.0005 sin(t) + 0.0838 x0\_tt  x0\_tt = 0.0061 1 + 0.0013 sin(t) + 1.1716 x0\_t + 0.0061 1 + -0.1388 x0  1 = 0.9999 1  x0 = 0.0306 1 + -0.0001 sin(t) + 0.0306 1 + 2.4187 x0^2  x0^2 = -0.0079 1 + -0.0079 1 + 0.2956 x0  x0^3 = -0.0022 1 + -0.0022 1 + 0.0708 x0  x0^4 = 0.0035 x0  x0^5 = 0.0000  x0^6 = 0.0000 | Model mirrors on the x axis? Not sure why  -Its the very last point – its an outlier (maybe due to being turned off?)  -Can see it a bit clearer with the scatter plot model (in blue)  Can see from 1 = 0.9999 1 that its not fully accurate |
| Ensemble-SINDy  (V1 and V2 only) | Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5']    Ensemble with replacement (V1)  (x0)' = -0.007 1 + 0.494 x0 + -10.244 x0^2 + 83.012 x0^3 + -265.069 x0^4 + 281.022 x0^5    Ensemble model without replacement (V2)  (x0)' = -0.007 1 + 0.494 x0 + -10.244 x0^2 + 83.012 x0^3 + -265.069 x0^4 + 281.022 x0^5 | V1 and V2 models are the exact same?   * There was a warning that the ensemble arguments are now deprecated – so may need to change code to see if its worked correctly/needs fixed |
| Note ensemble doesn’t work when I try to cut off the outliers (by cutting ‘x\_train’ short) | | |
| Ensemble-SINDy  (other versions) | Ensemble library (V3)  (x0)' = 0.249 x0 + -7.316 x0^2 + 67.718 x0^3 + -229.386 x0^4 + 250.693 x0^5    Both types of ensembling w/ candidate drops (V4)  (x0)' = -0.637 x0^5 | V3 is ensembling the library, V4 is ensembling both the data and the library, w/ candidate drops (see psindy>examples>13 for more detail)  Note: V4 changed when I reran the code – doesn’t fit the model well (at least 1-3 fit the majority of the data) |
| Ensemble-SINDy  (V1-4) | Ensemble with replacement (V1)  (x0)' = -0.007 1 + 0.494 x0 + -10.244 x0^2 + 83.012 x0^3 + -265.069 x0^4 + 281.022 x0^5    Ensemble model without replacement (V2)  (x0)' = -0.007 1 + 0.494 x0 + -10.244 x0^2 + 83.012 x0^3 + -265.069 x0^4 + 281.022 x0^5    Ensemble library (V3)  (x0)' = -0.007 1 + 0.494 x0 + -10.244 x0^2 + 83.012 x0^3 + -265.069 x0^4 + 281.022 x0^5    Both types of ensembling w/ candidate drops (V4)  (x0)' = 0.021 1 + -0.639 x0 + 5.980 x0^2 + -16.877 x0^3 + 32.817 x0^5 | V3 and V4 are different again   * Something must be random here   V1 and V2 are still the same, now V3 is also the same  V4 looks like the closest fit (at the start) |
| Ensemble-SINDy  (V1-4)  Threshold=0.05 (was 0.00) | Library:  ['1', 'x0', 'x0^2', 'x0^3', 'x0^4', 'x0^5', 'x0^6']    Ensemble with replacement (V1)  (x0)' = 0.074 x0^2 + -0.333 x0^3    Ensemble model without replacement (V2)  (x0)' = 0.074 x0^2 + -0.333 x0^3    Ensemble library (V3)  (x0)' = -0.087 x0 + 1.783 x0^2 + -8.910 x0^3 + 12.015 x0^4    Both types of ensembling w/ candidate drops (V4)  (x0)' = -0.096 x0^3 + -0.064 x0^4 |  |
|  |  |  |
|  |  |  |
|  |  |  |

SINDy PI –

Chart

Description automatically generated

1 = 1.0000 1 + 0.0001 x0 + -0.0005 x0^2 + 0.0011 x0^3 + 0.0003 x0^4

Chart

Description automatically generated with medium confidence

sin(t) = 0.0670 1 + -0.0001 sin(t) + 0.0001 sin(t/3) + 0.0011 sin(0.5t) + -0.0003 sin(2t) + 1.1088 x0\_t + 2.2887 x0\_tt + 0.0670 1 + -3.7720 x0 + 30.6186 x0^2 + -64.6056 x0^3 + -36.1895 x0^4 + -11.7446 x0^5 + -2.9029 x0^6

Chart

Description automatically generated

sin(t/3) = 0.2338 1 + 0.0009 sin(t/3) + 0.0036 sin(0.5t) + -0.0015 sin(2t) + -10.8341 x0\_t + -0.8716 x0\_tt + 0.2338 1 + -13.5187 x0 + 114.4788 x0^2 + -251.9807 x0^3 + -179.7410 x0^4 + -74.2854 x0^5 + -24.0877 x0^6

Chart

Description automatically generated

sin(0.5t) = 0.0961 1 + 0.0011 sin(t) + 0.0036 sin(t/3) + -0.0008 sin(2t) + -7.8091 x0\_t + 1.1462 x0\_tt + 0.0961 1 + -5.1807 x0 + 40.7867 x0^2 + -91.9282 x0^3 + -8.5401 x0^4 + 21.2035 x0^5 + 13.5667 x0^6

Chart

Description automatically generated

sin(2t) = -0.0629 1 + -0.0003 sin(t) + -0.0016 sin(t/3) + -0.0007 sin(0.5t) + -0.0001 sin(2t) + -7.5787 x0\_t + 4.9200 x0\_tt + -0.0629 1 + 5.0119 x0 + -68.2159 x0^2 + 359.8071 x0^3 + -486.9919 x0^4 + -489.6647 x0^5 + -271.5915 x0^6

Chart

Description automatically generated

x0\_t = 0.0021 1 + -0.0002 sin(t/3) + -0.0001 sin(0.5t) + -0.0001 sin(2t) + 0.0270 x0\_tt + 0.0021 1 + -0.1110 x0 + 0.7836 x0^2 + -0.7824 x0^3 + -0.8460 x0^4 + -0.4440 x0^5 + -0.1747 x0^6

Chart, line chart

Description automatically generated

x0\_tt = -0.0004 1 + 0.0003 sin(t) + -0.0001 sin(t/3) + 0.0002 sin(0.5t) + 0.0007 sin(2t) + 0.2852 x0\_t + -0.0004 1 + 0.0181 x0 + -0.0849 x0^2 + -0.0331 x0^3 + -0.0144 x0^4 + -0.0076 x0^5 + -0.0038 x0^6

Chart, line chart

Description automatically generated

1 = 1.0000 1 + 0.0001 x0 + -0.0005 x0^2 + 0.0011 x0^3 + 0.0003 x0^4

Chart

Description automatically generated with medium confidence

x0 = 0.0138 1 + 0.0200 x0\_t + 0.0048 x0\_tt + 0.0138 1 + 0.0002 x0 + 12.5113 x0^2 + -64.1203 x0^3 + 114.2033 x0^4 + 3.3844 x0^5 + -38.6638 x0^6

Chart, histogram

Description automatically generated

x0^2 = -0.0012 1 + -0.0020 x0\_t + -0.0005 x0\_tt + -0.0012 1 + 0.0832 x0 + 4.7234 x0^3 + -6.6159 x0^4 + -3.4030 x0^5 + -1.2141 x0^6

Chart, line chart

Description automatically generated

x0^3 = 0.0008 1 + -0.0015 x0\_t + 0.0008 1 + -0.0480 x0 + 0.4004 x0^2 + 0.1545 x0^4 + 0.0549 x0^5 + 0.0174 x0^6

Chart, line chart

Description automatically generated

x0^4 = 0.0004 1 + -0.0012 x0\_t + 0.0004 1 + -0.0178 x0 + 0.1067 x0^2 + 0.0555 x0^3 + 0.0063 x0^5 + 0.0019 x0^6

Chart, line chart

Description automatically generated

x0^5 = 0.0001 1 + -0.0005 x0\_t + 0.0001 1 + -0.0054 x0 + 0.0286 x0^2 + 0.0162 x0^3 + 0.0061 x0^4 + 0.0006 x0^6

Chart, line chart

Description automatically generated

x0^6 = -0.0002 x0\_t + -0.0015 x0 + 0.0076 x0^2 + 0.0045 x0^3 + 0.0018 x0^4 + 0.0006 x0^5

Chart, line chart

Description automatically generated

Conclusion: plotting x0 = … is the one that matches the raw data best

|  |  |
| --- | --- |
| Threshold |  |
| 1 |  |
| 0.5 |  |
| 0.4 |  |
| 0.3 |  |
| 0.2 |  |
| 0.1 |  |