



Auswahl und Optimierung historischer Daten zum Anlernen künstlicher Intelligenz in der kraftwerkstechnischen Verfahrenstechnik

Dipl.-Ing. Frank Gebhardt
NeuronalNetWorks! GmbH

 **NEURONALNETWORKS!**

Inhalt des heutigen Vortrages

- **Einleitung** in den Vortrag /Schwerpunkte
- **Herausforderungen** von „Big Data“
- **Theorie** zu Vorhersagen von Prozessgrößen mit NN
- Integrale **Vorhersage**
- **Auswahl** und **Bewertung** von Trainings- und Testdaten
- **Praxisbeispiel**

Einleitung in den Vortrag / Schwerpunkte

- Bei Anwendung von künstlicher Intelligenz (KI) in der Kraftwerks-Verfahrenstechnik
 - ▶ Begegnung mit „Big Data“!
- Verfahren zur Auswahl und Optimierung historischer Daten zum Anlernen von KI
- Anwendungen zur Vorhersage von Prozessgrößen:
 - Dampfproduktion
 - Temperaturkesseldecke
 - CO
 - etc...
- Hoher Zeitaufwand für Auswahl von Lernmustern
 - ▶ > 64.0000 Lernmuster entspricht > 1,5 Gbyte an Daten

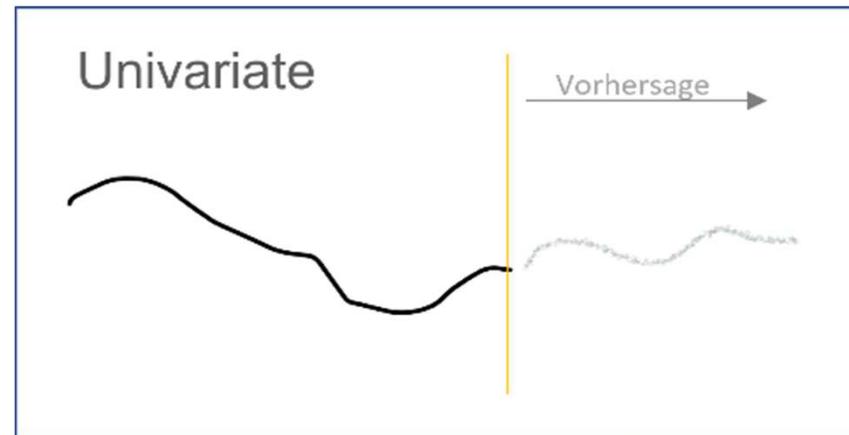
Herausforderungen von „Big Data“

- Datenmenge
 - kann nicht verarbeitet werden, da zum Beispiel bei Excel Obergrenzen bei der Zeilenanzahl bestehen
- Performance Probleme
 - gibt es bei z.B. bei Excel, wenn man eine Datei > 1 GByte öffnen will
 - bedeuten u.U. mehrere Minuten Wartezeit
- Suche nach Alternativen ist erforderlich
- Algorithmen für eine Vorverarbeitung der Trainings-Daten sehr hilfreich!
 - ▶ Professionelle Bearbeitung erforderlich für Wirtschaftlichkeit!

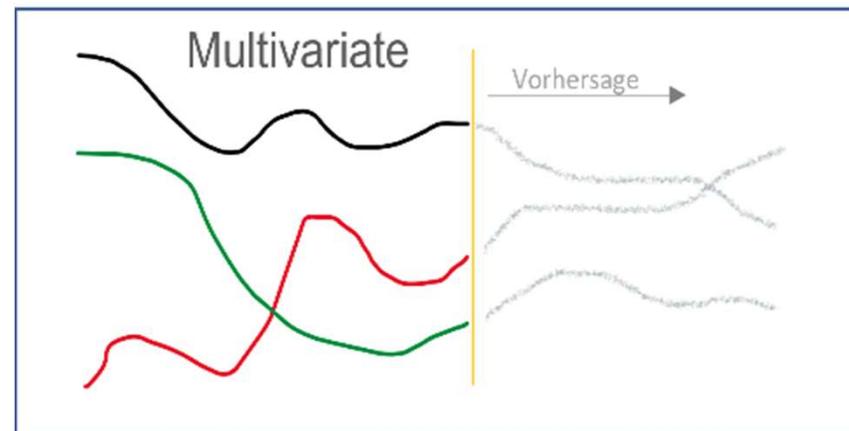
Theorie zu Vorhersagen von Prozessgrößen mit NN 1v2

Verschiedene Ansätze und Möglichkeiten für Vorhersage von Prozessgrößen:

- Abhängigkeit von nur **einer** Variablen:
 - Es wird nur ein Signal verwendet, um zukünftiges Verhalten vorherzusagen, z. B. nur Dampf
 - Diese Abhängigkeit wird in der Mathematik „univariat“ genannt



- Abhängigkeit von **mehreren** Variablen:
 - Es werden mehrere Signale verwendet, um zukünftiges Verhalten vorherzusagen, z.B. Dampf, Temperatur, CO, Luft, ...
 - Diese Abhängigkeit wird in der Mathematik „multivariat“ genannt

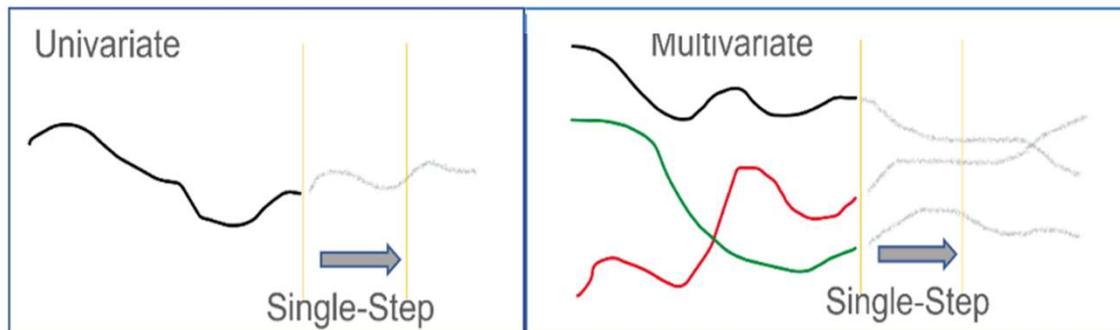


Theorie zu Vorhersagen von Prozessgrößen mit NN 2v2

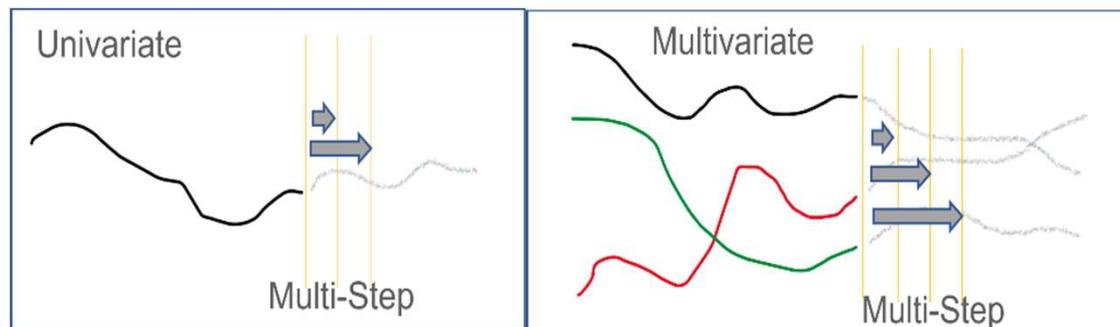
Prognosehorizonte

Es gibt verschiedene Prognosehorizonte für die Vorhersage von Prozessgrößen:

- Single-Step: nur ein Prognosehorizont, z.B. nur 5 Minuten:

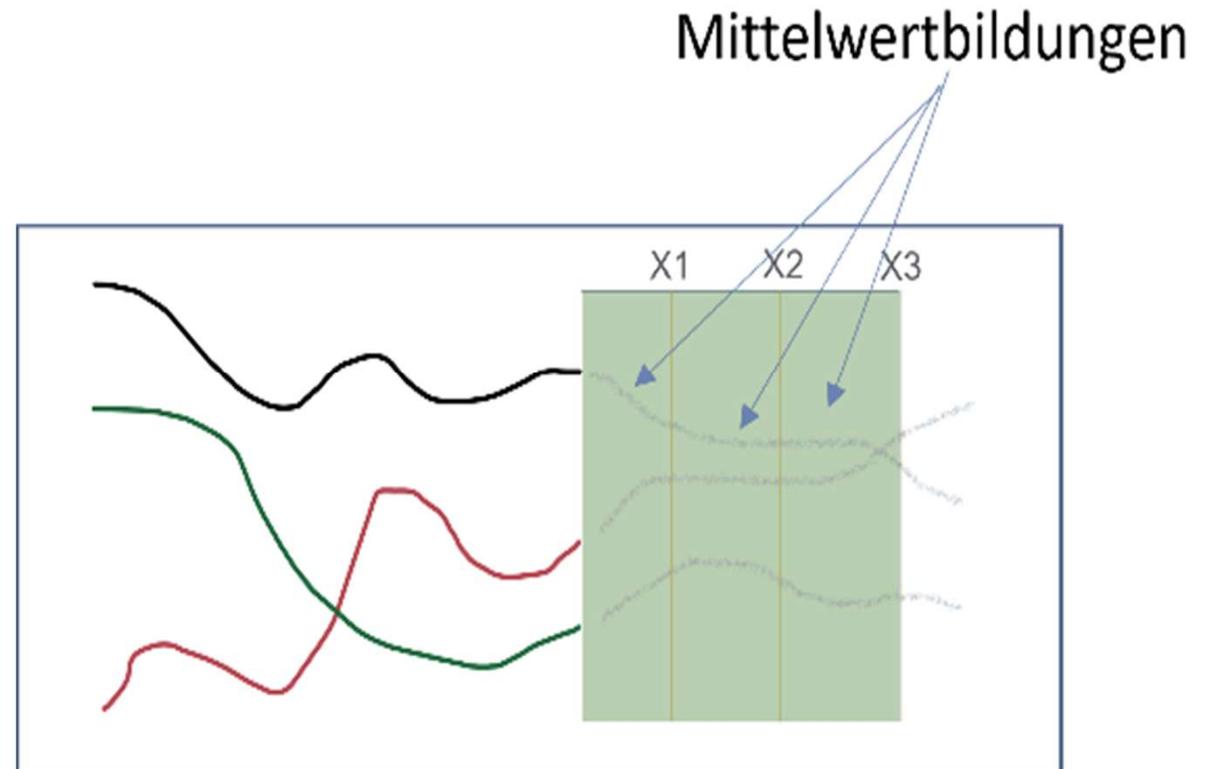


- Multi-Step: Vorhersage mehrerer Vorhersagehorizonte, z.B. 5 Minuten, 10 Minuten, ...



Integrale Vorhersage

- Multivariate und mehrstufige Vorhersagemethode
- Mittelung über Prognosehorizont
- **Mittelwerte:**
 - leichter vorherzusagen
 - enthalten die verfahrenstechnisch wichtigen Informationen
- Bestätigung durch die **Praxis**
 - ▶ Vorhersage von Dampfmassenströmen



Auswahl und Bewertung von Trainings- und Testdaten 1v5

- Trainings-Datensatz
 - 4 bis 6 Monaten für Vorhersagemodell, z.B. Dampfvorhersage
 - Bereinigung von Anlagenstillständen und störenden Prozessereignissen
 - Speicherung von so vielen Daten wie möglich
- Testdaten
 - Auswahl von kleinem Testset von wenigen Wochen
 - ▶ Das sind die Daten, die das neuronale Netz nicht gelernt hat, also nicht kennt
- Overfitting
 - Typische Strategien zur Vermeidung von Overfitting scheinen hier nicht zu funktionieren
 - Große Datensätze sind der Schlüssel!
 - Praktische Bestätigung bei Dampfvorhersage und Temperaturvorhersagen Kesseldecke
 - ▶ Verwendung sehr, sehr großer Datensätze erforderlich!
 - .

Auswahl und Bewertung von Trainings- und Testdaten 2v5

Komplexitätsoptimierung (bei gleicher NN-Topologie)

- **Großer** Datensatz

Bei **Unteranpassung** (schlechtes Lernen) ist die Komplexität des NN zu gering, um die Komplexität des Datensatzes zu modellieren

▶ Lösung: Anzahl der trainierbaren Ausgänge reduzieren, um die verfügbare Komplexität pro Ausgang zu erhöhen bzw. Variarität verkleinern

- **Kleiner** Datensatz

- Bei **Überanpassung** ist die Komplexität des NN zu groß

- Das NN erinnert sich an jede Stichprobe des Trainingsdatensatzes, die Verallgemeinerung von Testdaten wird in den meisten dieser Fälle schlecht sein

▶ Lösung: Anzahl der trainierbaren Ausgänge erhöhen oder Trainingsdataset vergrößern bzw. Variarität vergrößern

Auswahl und Bewertung von Trainings- und Testdaten 3v5

Ungleichmäßige Verteilung von Daten in einem Datensatz

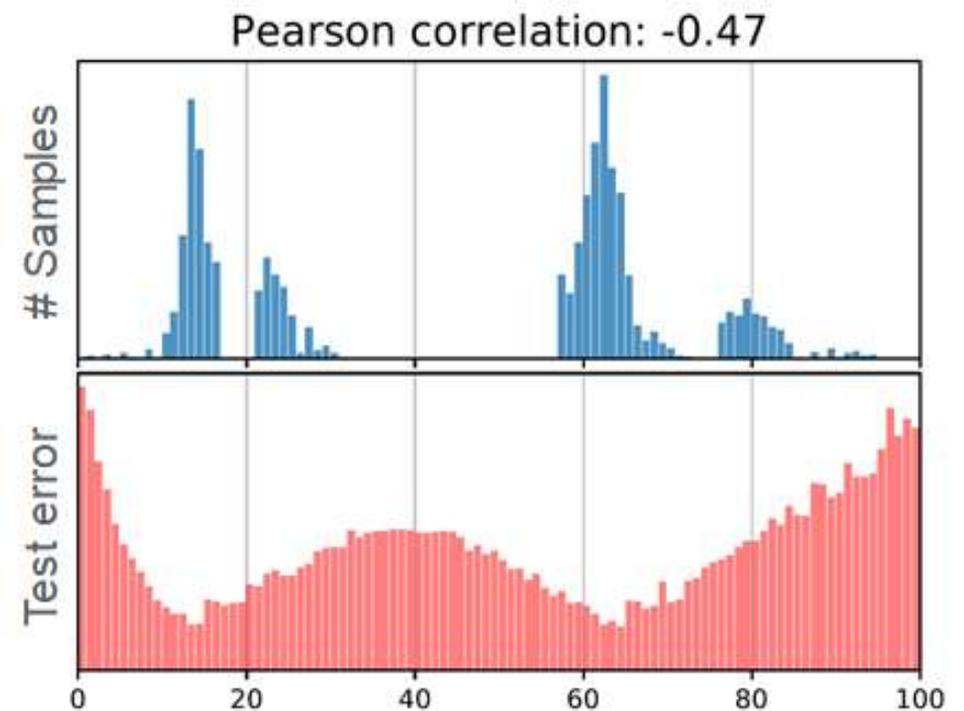
- „Unausgewogene Regression“
- Große 6-Monats-Datensätze spiegeln typische Verteilung der tatsächlichen Betriebsabläufe in der Anlage wieder

Anzahl Eingabe/Ausgabe-Paaren	Räumliche Dichte	Wahrscheinlichkeit für alternative ähnliche Eingabe/Ausgabe-Paare
Häufung	hoch	hoch
Mangel	niedrig	niedrig

Auswahl und Bewertung von Trainings- und Testdaten 4v5

Beispiel einer Datensatzverteilung

- ▶ Negative Korrelation (-0,47) zwischen der Anzahl der Stichproben für ein bestimmtes Betriebsverhalten und einem Prüffehler
- **Häufige** Daten in einem bestimmten Betriebsverhalten führen zu einem **geringen Testfehler** für dieses Betriebsverhalten
- **Seltene** Daten in einem bestimmten Betriebsverhalten führen zu einem **hohen Testfehler** für dieses Betriebsverhalten.

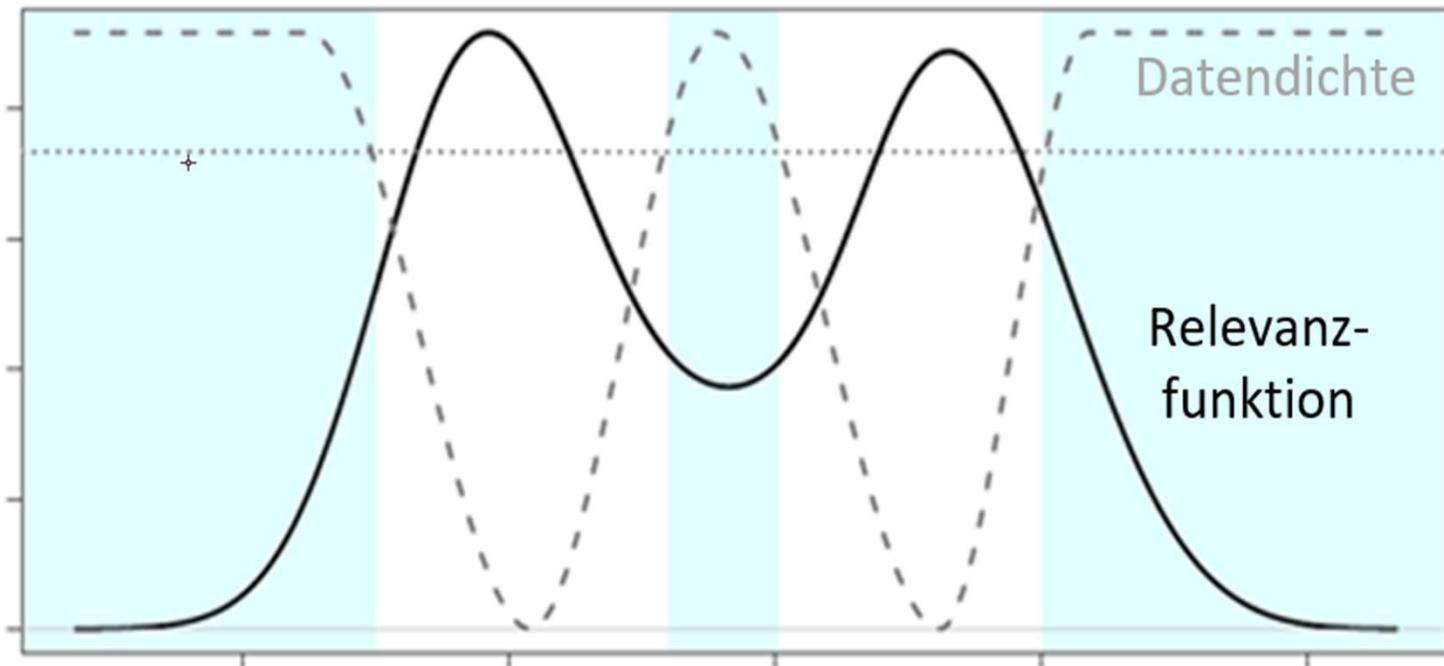


Auswahl und Bewertung von Trainings- und Testdaten 5v5

Relevanzfunktion

Für Verbesserung des Lernerfolges des NN:

- ▶ **Invertierte** Datendichte wird als Relevanzfunktion verwendet



Es gibt aus der Literatur viele theoretische Verfahren dazu:

- SMOTER (2013)
- SMOGN (2017)
- **WERCS (2018)**
- Dense Loss (2021)

Dampfvorhersage:

- Methode WERCS
- Spezifische Wichtung der Synapsengewichte
- Starke Verbesserung
- Praxistauglich!

Praxisbeispiel - Vorgehensweise

1v8

Standard-Vorgehensweise bei Balancing-Algorithmen zum Ausgleich von Lernmustern gemäß v.g. Methoden:

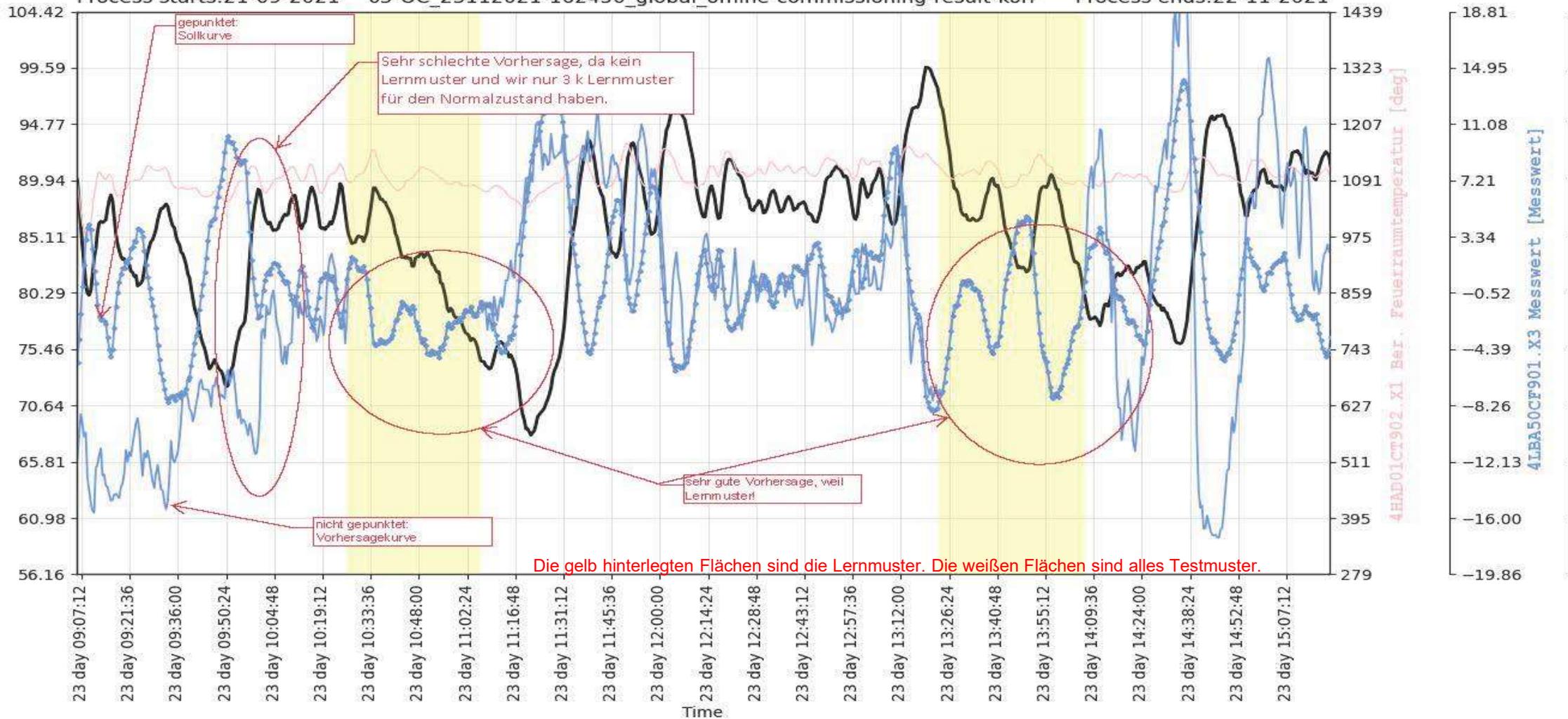
1. **Analyse** des kompletten Datensatzes auf häufige und seltene Daten
2. **Ausgleichung** der häufigen und seltenen Fälle, indem Sie eine Kombination aus:
 - **Oversampling** seltener Daten durch Erstellen synthetischer Daten, z. B. durch Hinzufügen von zufälligem Rauschen zu vorhandenen Daten
 - **Undersampling** für Datenhäufungen, indem Daten gefunden werden, die:
 - nahe beieinander liegen und daher annähernd gleich sind
 - zeitlich versetzt und ähnlich sindund deshalb gelöscht werden können.

Praxisbeispiel - Dampf-Vorhersage MVA

2v8

4k-Lernmuster-3k-Normal-1k-Dampfeinbruch

Process starts:21-09-2021 03-OC_23112021-162450_global_offline-commissioning-result-korr Process ends:22-11-2021

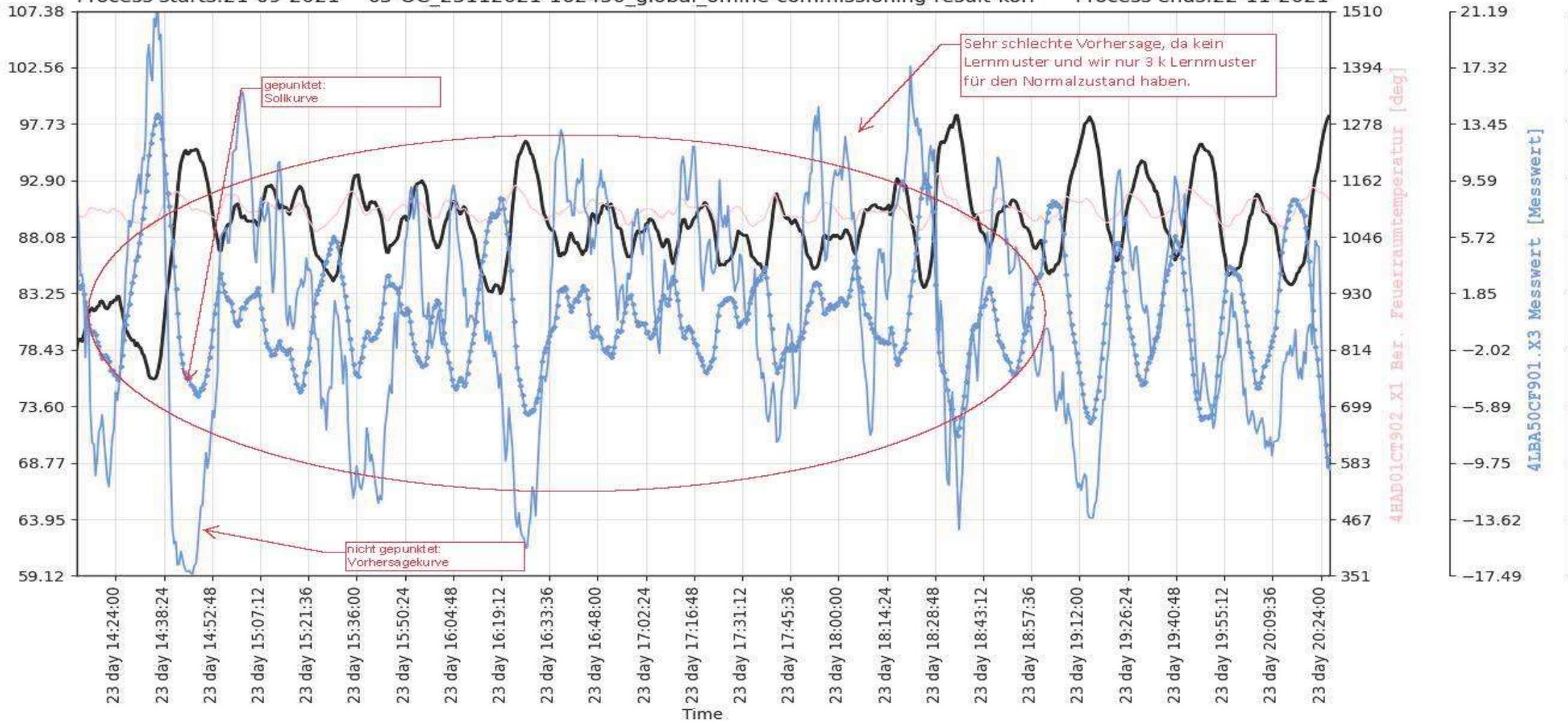


Praxisbeispiel - Dampf-Vorhersage MVA

3v8

Folgendes Testmuster ergab sich für den Normalzustand (kein Dampf-Einbruch)

Process starts:21-09-2021 03-OC_23112021-162450_global_offline-commissioning-result-korr Process ends:22-11-2021



Praxisbeispiel - Dampf-Vorhersage MVA

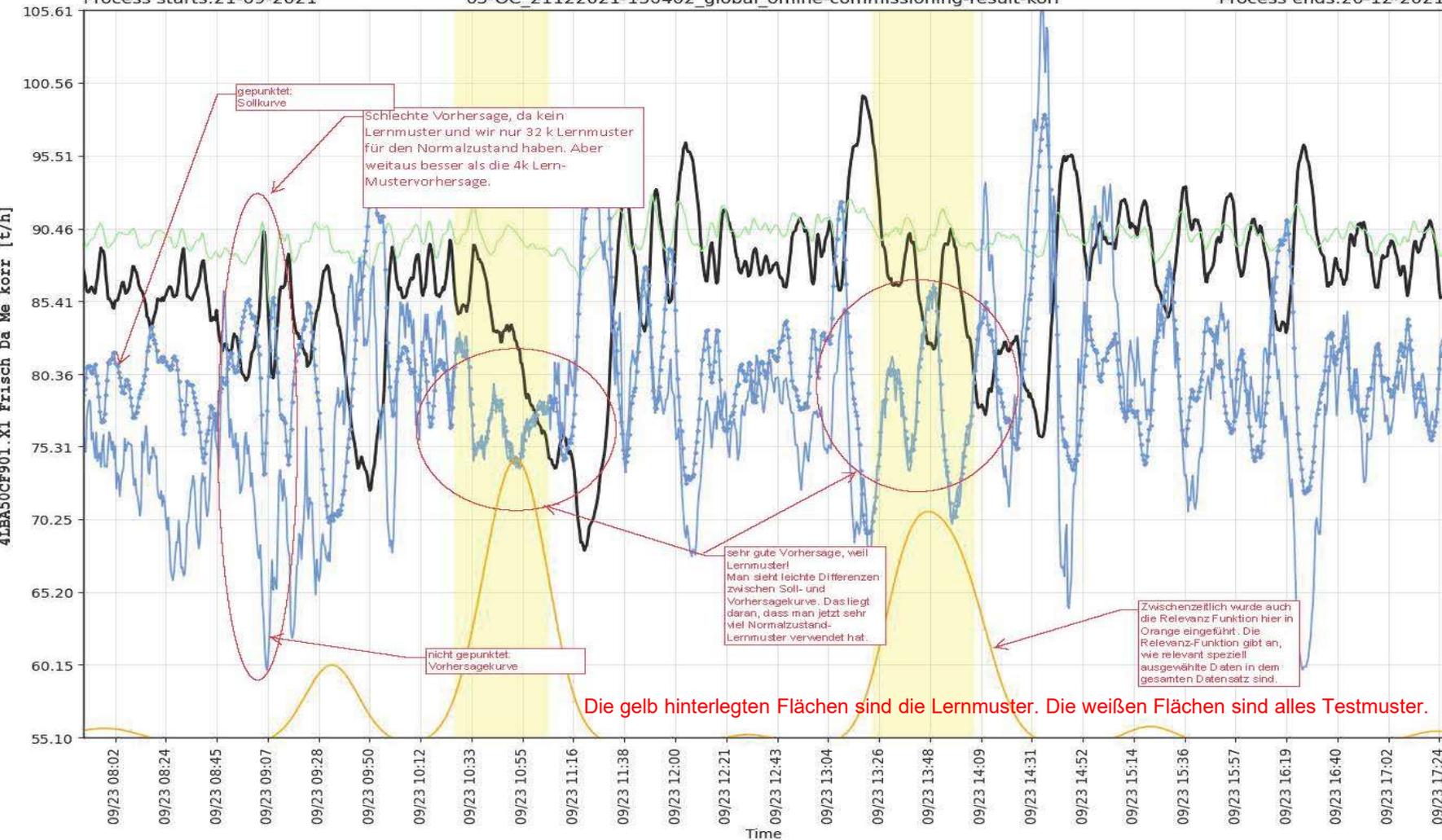
4v8

32k-Lernmuster-30k-Normal-2k-Dampfeinbruch

Process starts:21-09-2021

03-OC_21122021-130402_global_offline-commissioning-result-korr

Process ends:20-12-2021



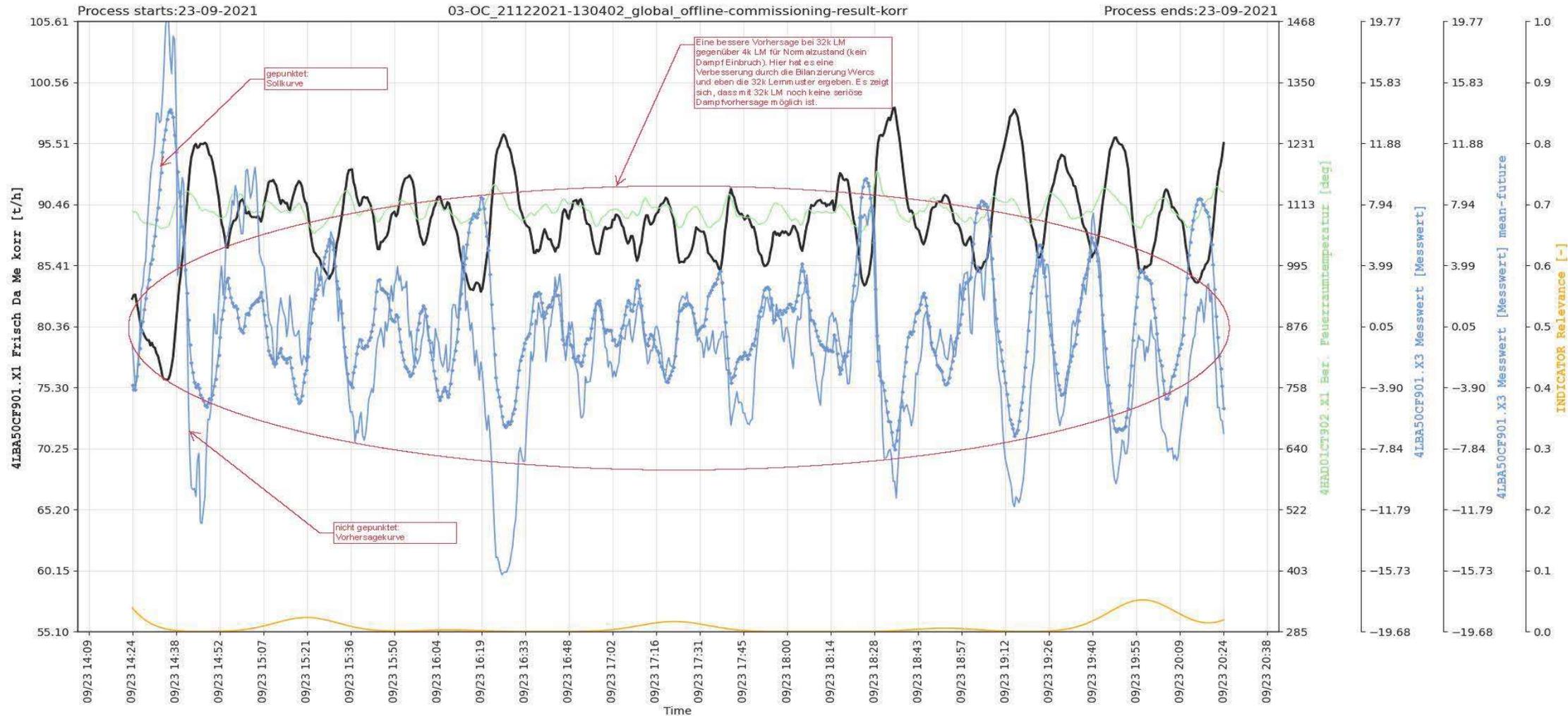
Relevanz Funktion:

- Trend in Orange
- Hohe Relevanzwerte führen zur Verstärkung von Lernmustern!

Praxisbeispiel - Dampf-Vorhersage MVA

5v8

Folgendes Testmuster ergab sich für den Normalzustand (kein Dampf-Einbruch)

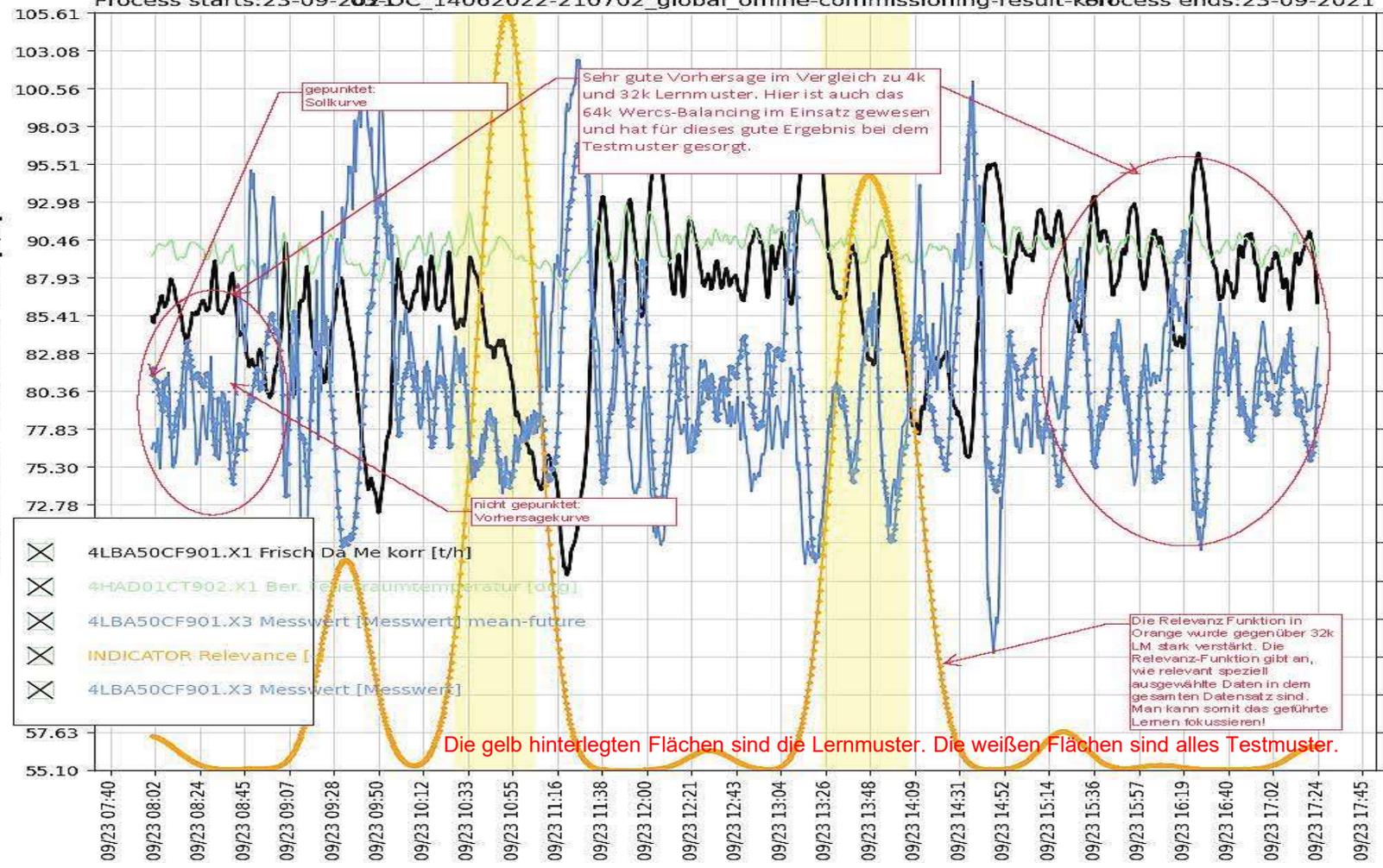


Praxisbeispiel - Dampf-Vorhersage MVA

6v8

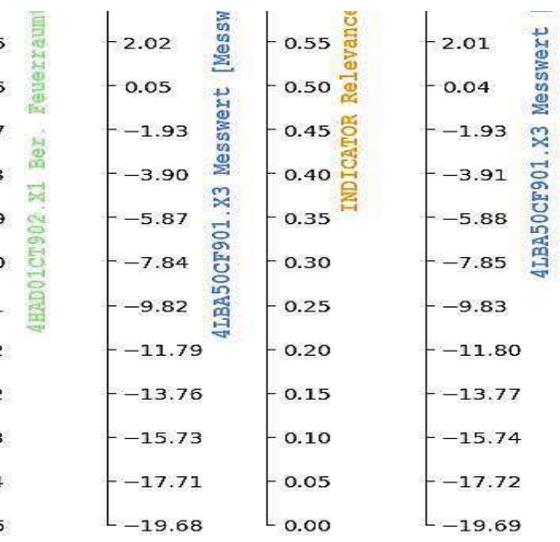
64k-Lernmuster-62k-Normal-2k-Dampfeinbruch

Process starts:23-09-2021 DC_14062022-210702_global_offline-commissioning-result-kö process ends:23-09-2021



Die gelb hinterlegten Flächen sind die Lernmuster. Die weißen Flächen sind alles Testmuster.

- Auswahl der Lernmuster (LM)
- „2k-Dampfeinbruch“: manuell
 - 62k-LM: Werchs-Bilanzierung und Relevanz-Verstärkung

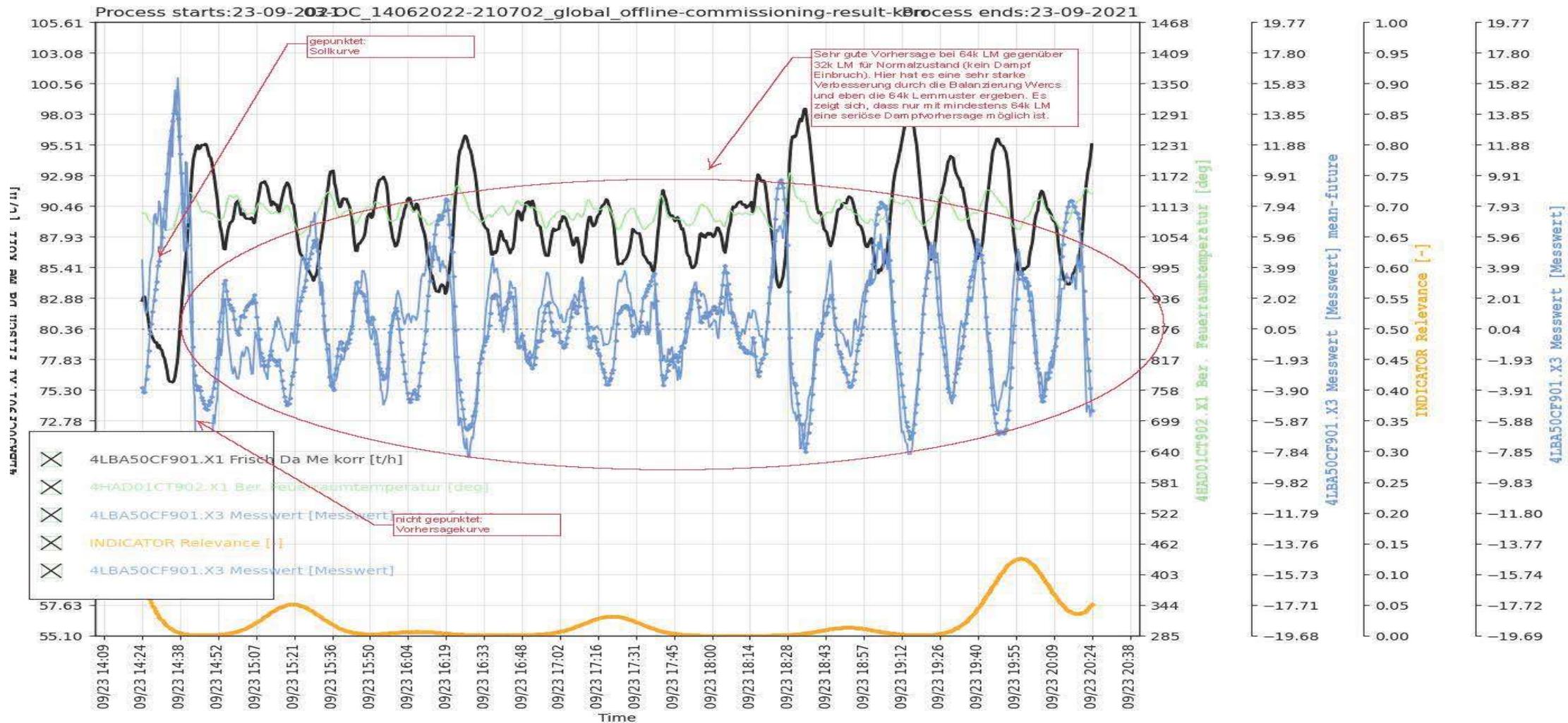


×	4LBA50CF901.X1 Frisch Da Me korr [t/h]
×	4HAD01CT902.X1 Ber. Feuerraumtemperatur [00g]
×	4LBA50CF901.X3 Messwert [Messwert] mean-future
×	INDICATOR Relevance [
×	4LBA50CF901.X3 Messwert [Messwert]

Praxisbeispiel - Dampf-Vorhersage MVA

7v8

Folgendes Testmuster ergab sich für den Normalzustand (kein Dampf-Einbruch):



Praxisbeispiel - Dampf-Vorhersage MVA

8v8

Relevanz / verfahrenstechnischer Indikator – Definition

- Die Relevanz wurde auch als verfahrenstechnischer Indikator verwendet, der hier „Prozesskategorie 1: Überschüttung Müllrost“ genannt wurde (Doppelfunktion!)
- Trends von Messungen:
 - Primärluftdrücke: Trend ansteigend
 - Rauchgas O₂-Gehalt: Trend ansteigend
 - FLR-Regelung-Rost-Vorschub: Trend ansteigend
 - Feuerraumtemperatur: Trend abfallend
 - Feuchtemessung: Trend ansteigend
- Zukunftswerte der Dampfproduktion
 - 5 Minuten: Trend abfallend
 - 15 Minuten: Trend abfallend
 - 30 Minuten: Trend abfallend
- Berechnung Relevanz / verfahrenstechnischer Indikator
 - Mathematische Verknüpfung von:
 - Trends von Messungen
 - Zukunftswerte der Dampfproduktion



NEURONALNETWORKS!

Frank Gebhardt

Geschäftsführer

Frank.Gebhardt@neuronalnetworks.de

+49 177 611 5983

