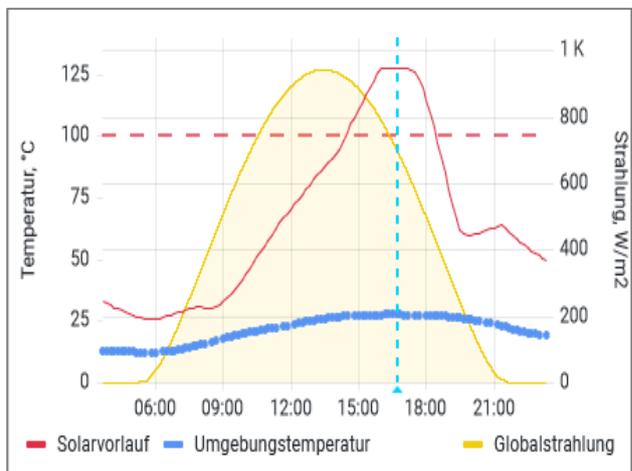




Abschlussbericht vom 10.04.2024

# LoCoSol+

## Low-Cost Monitoring thermischer Solaranlagen mit IoT-Sensor und maschinellem Lernen





**Datum: 10.04.2024**

**Ort:** Bern

**Subventionsgeberin:**

Bundesamt für Energie BFE  
Sektion Energieforschung und Cleantech  
CH-3003 Bern  
[www.bfe.admin.ch](http://www.bfe.admin.ch)

**Subventionsempfänger/innen:**

Fachhochschule Nordwestschweiz, FHNW  
Hofackerstrasse 30, CH-4132 Muttenz  
<https://www.fhnw.ch/de/die-fhnw/hochschulen/architektur-bau-geomatik/institute/ineb>

Energie Zukunft Schweiz AG  
Viaduktstrasse 8, 4051 Basel  
<https://energiezukunftschweiz.ch/>

**Autor/in:**

Parantapa Sawant, FHNW, [parantapa.sawant@fhnw.ch](mailto:parantapa.sawant@fhnw.ch)  
Barbara Sintzel, FHNW, [barbara.sintzel@fhnw.ch](mailto:barbara.sintzel@fhnw.ch)  
Ralph Eismann, FHNW, [ralph.eismann@fhnw.ch](mailto:ralph.eismann@fhnw.ch)  
Joachim Hofmann, FHNW

Bernd Sitzmann, [b.sitzmann@energiezukunftschweiz.ch](mailto:b.sitzmann@energiezukunftschweiz.ch)

**BFE-Projektbegleitung:**

Dr. Stephan Mathez, [stephan.a.mathez@solarcampus.ch](mailto:stephan.a.mathez@solarcampus.ch)  
Andreas Eckmanns, [andreas.eckmanns@bfe.admin.ch](mailto:andreas.eckmanns@bfe.admin.ch)

**BFE-Vertragsnummer:** SI/502318-01

**Für den Inhalt und die Schlussfolgerungen sind ausschliesslich die Autoren dieses Berichts verantwortlich.**



## Zusammenfassung

Das Projekt behandelte die kritische Notwendigkeit, die Überwachung von solarthermischen Anlagen in Bezug auf einfache Implementierung, Zuverlässigkeit und Kosteneffizienz zu verbessern. Dies wurde erreicht durch minimalen Hardwareaufwand in Form eines einzigen LoRaWAN Sensor zur Messung der Vorlauftemperatur, einer cloudbasierten Datenbank mit automatischer Einbindung von Wetterdaten der nächstgelegenen Wetterstation und eines regelbasierten Algorithmus (RBA) zur automatisierten Datenanalyse und Generation von Fehlermeldungen. Darüber hinaus waren alle Software-Tools Open Source. Die Grundlagen wurden bereits in einem früheren Projekt, «LoCoSol», gelegt. Ziel des aktuellen Projekts war es, die Genauigkeit der Fehlererkennung zu erhöhen, den Solarertrag zu schätzen und den Bewertungsprozess in einem benutzerfreundlichen und robusten Rahmen vollständig zu automatisieren.

Die quantitative und qualitative Analyse der Ergebnisse bestätigte, dass die angestrebte Genauigkeit für die verschiedenen Key Performance Indikatoren (KPIs) erreicht wurde, z.B. 98% für den Pumpenbetrieb und 93% für die Stagnationserkennung. Zusätzlich wurden vier neue KPIs für die Erkennung von Wärmeverlusten durch Nachtauskühlung mit einer zufriedenstellenden Genauigkeit von mindestens 86% implementiert. Der neue RBA schätzt auch den zu erwartenden Solarertrag für jede Anlage ausreichend ab. Gegenüber dem im Vorgängerprojekt entwickelten hybriden Algorithmus wurden mit dem RBA bereits gleichwertige, beziehungsweise für einige KPIs sogar höhere Genauigkeiten erreicht.

Im Hinblick auf das Software-Framework wurde ein automatisiertes Test-Setup erstellt, welches die Zuverlässigkeit des Codes erhöht, und die RBA wurde in die Datenbankumgebung des Unternehmens implementiert. Dies erhöhte die Robustheit, da die Anzahl der beteiligten Softwaretools im Vergleich zum Vorgängerprojekt reduziert wurde. Die erfolgreiche Implementierung eines Überwachungssystems zur Erkennung von Betriebsfehlern oder Abweichungen vom geplanten Solarertrag aufgrund von Konstruktions- und Installationsfehlern an 468 Anlagen (in einem relativ kurzen Zeitraum) hat die Skalierbarkeit und Kosteneffizienz dieses Frameworks bewiesen. Im Vergleich zu einem typischen drahtgebundenen Datenerfassungssystem mit TCP/IP-Kommunikationsprotokoll und mehreren Sensoren zur Überwachung einer Anlage sind geringere Kosten zu erwarten.

Eine kritische Diskussion der Ergebnisse zeigt jedoch auch die Grenzen der RBA in Bezug auf die Qualität der verwendeten Daten und den hohen Aufwand für die Anpassung der Parameter. Hier untersucht das Projekt die Integration von Algorithmen des maschinellen Lernens zur automatischen Fehlererkennung und zur verbesserten Schätzung der in der RBA verwendeten Parameter. Diese Methoden haben sich als vielversprechend erwiesen, wenn sie in Kombination mit der RBA eingesetzt werden, um ein automatisiertes und genaueres System zu erhalten.



## Summary

The project addressed the critical need to improve the monitoring of solar thermal systems in terms of ease of implementation, reliability, and cost efficiency. This was achieved through minimal hardware effort in the form of a single LoRaWAN sensor to measure the outlet temperature, a cloud-based database with automatic integration of weather data from the nearest weather station and a rule-based algorithm (RBA) for automated data analysis and generation of error messages. In addition, all software tools were open source. The foundations had already been laid in an earlier project, "LoCoSol". The aim of the current project was to increase the accuracy of fault detection, estimate the solar yield and fully automate the evaluation process in a user-friendly and robust framework.

The quantitative and qualitative analysis of the results confirmed that the targeted accuracy for the various key performance indicators (KPIs) was achieved, e.g. 98% for pump operation and 93% for stagnation detection. In addition, four new KPIs were implemented for the detection of thermal losses due to night cooling with a satisfactory accuracy of at least 86%. The new RBA also sufficiently estimates the expected solar yield for each system. Compared to the hybrid algorithm developed in the previous project, the RBA has already achieved equivalent or even higher accuracies for some KPIs.

Regarding the software framework, an automated test setup was created to increase the reliability of the code, and the RBA was implemented in the company's database environment. This increased robustness as the number of software tools involved was reduced compared to the previous project. The successful implementation of a monitoring system to detect operational faults or deviations from the planned solar yield due to design and installation errors on 468 installations (in a relatively short period of time) has proven the scalability and cost-effectiveness of this framework. Compared to a typical wired data acquisition system with TCP/IP communication protocol and multiple sensors to monitor a system, lower costs can be expected.

However, a critical discussion of the results also shows the limitations of the RBA in terms of the quality of the data used and the high effort required to adapt the parameters. Here, the project is investigating the integration of machine learning algorithms for automatic fault detection and improved estimation of the parameters used in the RBA. These methods have proven to be promising when used in combination with RBA to obtain a more automated and accurate system.



## Résumé

Le projet visait à améliorer la surveillance des systèmes solaires thermiques en termes de facilité de mise en œuvre, de fiabilité et de rentabilité. Il s'agissait notamment d'utiliser un seul capteur LoRaWAN pour la température de sortie, d'intégrer les données météorologiques de la station la plus proche dans une base de données intégrée au cloud, et d'utiliser un algorithme basé sur des règles (RBA) pour l'analyse automatisée et le signalement des erreurs. L'utilisation d'outils logiciels libres et les enseignements tirés du projet LoCoSol ont permis de poser les bases du projet.

Les objectifs quantifiables du projet étaient d'améliorer la précision de la détection des erreurs, d'estimer le rendement solaire et d'automatiser entièrement le processus d'évaluation. Les résultats ont montré que les objectifs de précision ont été atteints pour divers indicateurs clés de performance (ICP), par exemple 98 % pour le fonctionnement de la pompe et 93 % pour la détection de la stagnation. Les nouveaux ICP pour la détection du mélange forcé dans le réservoir de stockage ont atteint une précision d'au moins 86 %. Le RBA a estimé efficacement le rendement solaire, la plupart des ICP surpassant les résultats de l'algorithme hybride du projet prédécesseur.

Le cadre logiciel comprenait une configuration de test automatisée pour la fiabilité du code, et le RBA a été mis en œuvre dans la base de données de l'entreprise, ce qui en a amélioré la robustesse. L'évolutivité et la rentabilité du cadre ont été prouvées par la mise en œuvre d'un système de surveillance dans 468 installations. Par rapport aux systèmes câblés traditionnels, on s'attend à une baisse des coûts.

Cependant, le projet a reconnu les limites du RBA, ce qui a conduit à l'exploration d'algorithmes d'apprentissage automatique pour la détection automatique des défauts et l'amélioration de l'estimation des paramètres. La combinaison de ces méthodes avec le RBA s'est révélée prometteuse pour un système plus automatisé et plus précis.



## Take-home messages

- Die Kombination von IoT, frei verfügbaren Wetterdaten und automatisierter Datenanalyse in einem Monitoringsystem ermöglicht:
  - o Eine einfache und kostengünstige Nachrüstung zur Optimierung des Betriebs bestehender Anlagen.
  - o Gezielte Serviceeinsätze bei auftretenden Störungen.
  - o Eine verbesserte Funktionsübersicht, die zu mehr Transparenz führt.
- Regelbasierte Algorithmen eignen sich gut für die Analyse täglicher Datensätze und können Genauigkeiten von mehr als 95% erreichen. Ihre Entwicklung erfordert jedoch eine manuelle Analyse des Systems, die bei einer grossen Anzahl von Systemen zeitaufwändig und fehleranfällig sein kann und umfangreiche Fachkenntnisse erfordert.
- Ein datengesteuerter maschineller Lernalgorithmus, der mit minimalem Konfigurationsaufwand auf eine Vielzahl von Systemlayouts angewendet werden kann und automatisch korrelierende Sensoren und Modelle identifiziert, ist von Vorteil.
- Die Wahl eines Softwaretools kann den Projekterfolg erheblich beeinflussen. Daher sollte ein Abgleich mit den Projektanforderungen, der Kompatibilität, den Funktionen und den Kompetenzen vorgenommen werden.



# Inhaltsverzeichnis

Zusammenfassung.....	3
Summary .....	4
Résumé.....	5
Take-home messages .....	6
Inhaltsverzeichnis .....	7
Abkürzungsverzeichnis .....	9
<b>1 Einleitung .....</b>	<b>10</b>
1.1 Ausgangslage und Hintergrund .....	10
1.2 Motivation des Projektes .....	10
1.3 Projektziele .....	11
1.3.1 Erhöhen der Treffsicherheit der Fehlererkennung .....	11
1.3.2 Vollständige Automatisierung der Auswertung.....	11
1.3.3 Ermitteln einer geeigneten Disseminationsstrategie .....	12
<b>2 Vorgehen und Methode.....</b>	<b>12</b>
2.1 Datenquellen für das Monitoring.....	12
2.1.1 Instrumentierung von Anlagen mit dem Temperatursensor .....	12
2.1.2 Wetterdaten .....	13
2.2 Produktivsystem und Entwicklungssystem.....	13
2.3 Datensätze für die Entwicklung der Algorithmen.....	15
<b>3 Entwicklung des verbesserten regelbasierten Algorithmus.....</b>	<b>16</b>
3.1 Datenvorbereitung .....	16
3.2 KPI-Definition.....	17
3.2.1 Logik für N1 (ersten signifikanten Temperaturanstiegs) .....	20
3.2.2 Logik für N2 (Pumpenstart) .....	20
3.2.3 Logik für N3 (Pumpenstopp).....	22
3.2.4 Logik für N4 (max. Vorlauftemperatur) .....	24
3.2.5 Logik für N5 (Betrieb ohne Nachtauskühlung) .....	24
3.2.6 Logik für N6 (Nachtauskühlung nach Aussen) .....	25
3.2.7 Logik für N7 (Rohrinterne Zirkulation) .....	26
3.3 Explorative Datenanalyse zur Parameterabschätzung .....	26
3.4 Testing .....	30
3.4.1 Ergebnisse der KPI-Analyse für die KPIs N1...N4.....	31
3.4.2 Ergebnisse der KPI-Analyse für die KPIs N5...N7.....	32
3.5 Labels und Attributes .....	34
3.6 Alarms.....	35
3.7 Implementierung.....	37



3.7.1	Implementierung in «Node-Red».....	37
3.7.2	Implementierung des RBA in «Flux» .....	39
3.8	Bewertung der Ergebnisse .....	41
<b>4</b>	<b>Einsatz maschinellen Lernens .....</b>	<b>42</b>
4.1	Zeitreihenanalyse .....	42
4.2	Exponentielle Glättung .....	44
4.3	SARIMAX.....	45
4.4	Bewertung der Ergebnisse .....	46
4.5	Empfehlung für die Softwarestruktur zur Implementierung der Zeitreihenmethode.....	46
<b>5</b>	<b>Ausblick.....</b>	<b>47</b>
<b>6</b>	<b>Literaturverzeichnis .....</b>	<b>48</b>



## Abkürzungsverzeichnis

EZS	Energie Zukunft Schweiz AG
IoT	Internet of Things
KPI	Key Performance Indikator
MLA	Machine Learning Algorithmus
LoRaWAN	Long Range Wide Area Network
RBA	Regelbasierter Algorithmus



# 1 Einleitung

## 1.1 Ausgangslage und Hintergrund

Solarthermische Anlagen müssen zur Sicherstellung des einwandfreien Betriebs, auf ihre Funktion überwacht werden. Die Überwachung dient nicht nur zur Messung und Dokumentation des Energieertrages. Noch wichtiger ist, dass Fehlfunktionen frühzeitig erkannt und die Ursachen zuverlässig ermittelt werden können. Die Notwendigkeit eines Monitorings zur Fehleridentifikation, analog dem gängigen Monitoring bei Photovoltaikanlagen wurde in zahlreichen Studien [1-7] eindrucksvoll belegt.

Auf der Grundlage dieser Studien sowie mit Berücksichtigung verschiedener Ansätze zur Fernüberwachung von solarthermischen Anlagen [8], wurde durch die Energie Zukunft Schweiz AG (EZS) ein neues, vergleichsweise kostengünstiges Monitoringsystem entwickelt und wird erfolgreich eingesetzt. Schlüsselkomponente des Systems ist ein IoT-Sensor, der die Vorlauftemperatur des Solarkreises misst. Der Sensor überträgt die Daten über das Long Range Wide Area Network (LoRaWAN) an eine Datenbank. Dem Datensatz werden die Umgebungstemperatur und die globale Bestrahlungsstärke der nächstgelegenen Wetterstation zugeordnet. Die Beurteilung des Betriebsverhaltens einer Solaranlage geschieht anhand von Key Performance Indikatoren (KPI), mit denen charakteristische Ereignisse und Zustände quantifiziert werden.

Dieses Monitoringsystem wurde im vorangegangenen Projekt [9] wissenschaftlich untersucht und ein Konzept für die Umsetzung in Echtzeit wurde erstellt. Die Tagesdatensätze aller Anlagen werden täglich durch einen regelbasierten Algorithmus (RBA) dahingehend analysiert, ob einer oder mehrere KPIs zutreffen. Aus den Resultaten dieser Analyse werden Fehlfunktionen abgeleitet (wiederholte Stagnation, Wärmeverlust durch Nachtauskühlung) und per E-Mail an EZS übermittelt. Einmalige Fehlfunktionen werden als spezielles Ereignis registriert, aber nicht als Fehler interpretiert.

Es hat sich gezeigt, dass der im vorangegangenen Projekt entwickelte RBA nicht in der Lage ist, zulässige von fehlerbedingten Stagnationseignissen zu unterscheiden. Ausserdem konnte die fehlerbedingte Nachtauskühlung nur dann detektiert werden, wenn der Sensor am Kollektorausstritt montiert war.

Aufgrund der Unzulänglichkeit des RBA wurde in dem vorangegangenen Projekt [9] auch die Möglichkeiten zur Implementierung von «Maschine Learning Algorithmen» (MLA) für die Parameterabstimmung identifiziert und getestet [10]. Allerdings standen für diesen Zweck bei Weitem zu wenig Datensätzen zur Verfügung. Aufgrund der zu geringen Datenmenge wurde die gewünschte Treffsicherheit nicht erreicht, und der MLA wurde nicht in das Monitoringsystem von EZS implementiert.

Es wurde jedoch nachgewiesen, dass die Beurteilung einer solarthermischen Anlage auf der Grundlage der Daten eines einzigen Temperatursensors grundsätzlich möglich ist. Das Verfahren wird seither durch EZS zur Überwachung einer stetig wachsenden Zahl von Solaranlagen<sup>1</sup> eingesetzt.

## 1.2 Motivation des Projektes

Das harmonisierte Fördermodell der Kantone [11] schreibt vor, dass solarthermische Anlagen mit über 20kW Leistung mit einem aktiven Monitoring ausgerüstet werden müssen, wobei auf die Massnahmen zum Qualitätsmanagement von Swissolar [12] verwiesen wird. Die Überprüfung des Zustandes geförderter Solaranlagen [1-5] durch EZS zeigt jedoch, dass diese Vorgaben praktisch nicht umgesetzt werden, obwohl Lösungen zur Fernüberwachung am Markt verfügbar sind. Diese Lösungen sind jedoch recht kostspielig und erfordern einen stabilen Internetzugang. Ausserdem ist der Installationsaufwand nicht zu unterschätzen, da in der Regel mehrere Sensoren installiert werden müssen.

---

<sup>1</sup> Stand 24.02.2024: 468 Anlagen



EZS hat gezeigt, dass die im vorangegangenen Projekt entwickelte Lösung umsetzbar ist und am Markt etabliert werden kann. Um die Verbreitung dieser eleganten und kostengünstigen Lösung nachhaltig zu sichern, musste der Algorithmus in zwei Richtungen verbessert werden. Erstens musste die Datenanalyse auf weitere Fehlerbilder erweitert werden. Zweitens musste die Treffsicherheit des Algorithmus erhöht werden. Dies ist nach den Resultaten des vorangegangenen Projekts möglich. Entsprechende Entwicklungspfade wurden erarbeitet und haben schliesslich zu diesem Forschungsprojekt geführt.

### 1.3 Projektziele

Hauptziel dieses Projekts war, der Solarthermiebranche eine ausgereifte Technologie für eine kostengünstige Funktionsüberwachung für thermische Solaranlagen zur Verfügung zu stellen. Diese Funktionsüberwachung zeichnet sich durch folgende Eigenschaften aus:

- Minimale Installationskosten durch Beschränkung auf einen einzigen Temperatursensor
- Installation des Sensors wahlweise im Technikraum oder beim Kollektorfeld
- Automatische Fehlererkennung und Meldung an den technischen Verantwortlichen der Anlage
- Periodische Übermittlung positiver Meldungen (Laufzeiten, Ertragsabschätzung)
- Schweizweit, flächendeckend einsetzbar
- Für Neuanlagen und Bestandsanlagen gleichermaßen geeignet.

Damit werden die Voraussetzungen für einen zuverlässigen Betrieb solarthermischer Anlagen geschaffen und ein Beitrag zur Förderung der Branche geleistet. Durch diese Funktionsüberwachung wird ausserdem erreicht, dass Ausfallzeiten kurz sind, und die Solaranlagen den geplanten Ertrag bringen. Auf der Grundlage der bereits geleisteten Vorarbeiten werden aus dem Hauptziel zwei technische Teilziele und ein sozioökonomisches Teilziel abgeleitet:

#### 1.3.1 Erhöhen der Treffsicherheit der Fehlererkennung

Seit Abschluss des vorangegangenen Projektes Mitte 2021 wurden rund 200 weitere Anlagen mit Sensoren ausgerüstet und in das Monitoring integriert. Diese neu verfügbaren Daten bilden zusammen mit dem bereits verwendeten Datensatz und weitere Felderfahrung eine weitaus umfassendere Basis für Verbesserung des bestehenden RBA. Der RBA wird dahingehend optimiert, dass in der Menge als fehlerhaft identifizierten Anlagen alle tatsächlich fehlerhaften Anlagen enthalten sind. Das Projekt zielt darauf ab, bestimmte Genauigkeitsstufen in zwei Hauptkategorien zu erreichen, erstens die Identifizierung von Fehlern und zweitens den Anlagenstatus. In der ersten Kategorie, bei der Bestimmung der «Nachtauskühlung des Speichers», ist das angestrebte Genauigkeitsniveau auf 95% festgelegt. Im Fall von «Stagnation aufgrund von Fehlfunktionen» wird eine Genauigkeit von 90% angestrebt. Bei «Konkurrierende Zusatzheizung» schliesslich wird ein Genauigkeitsgrad von 90% angestrebt. In der zweiten Kategorie ist die Identifizierung des Zustands 'Pumpe ein/aus' erforderlich und die Zielgenauigkeit beträgt 90%.

#### 1.3.2 Vollständige Automatisierung der Auswertung

Das Ziel besteht darin, einen Analyseprozess zu entwickeln, welcher es EZS ermöglicht, tägliche Auswertungen des Betriebsverhaltens von hunderten Solaranlagen automatisiert durchzuführen. Dabei werden verschiedene Software-Tools eingesetzt und verglichen. Das gewonnene Know-how soll EZS bei der Entscheidung für ein geeignetes System unterstützen, das auch die vorhandenen Kompetenzen und die Nutzeranforderungen berücksichtigt.



### 1.3.3 Ermitteln einer geeigneten Disseminationsstrategie

Es wird eine Strategie erarbeitet, mit welcher das Monitoring möglichst rasch und flächendeckend eingesetzt werden kann. Diese Strategie beinhaltet die geeignete Kommunikation und Wertedefinition mit folgenden Inhalten und Zielen:

- Monitoring soll von Anbietern, Investoren und Betreibern als wertsteigernde Innovation aufgefasst werden.
- Monitoring ist Ausdruck der Professionalität der Branche, die in diesem Punkt mit der Photovoltaik gleichzieht.
- Das Image der Solarthermie-Branche soll einen Aufschwung erfahren und von den modernen Werkzeugen der Digitalisierung und Energieinformatik profitieren.

Dies sind die Voraussetzungen dafür, dass sich der Solarthermie-Markt erholt und dadurch weitere qualitätssteigernde, kostensenkende Innovationen möglich werden.

Als Betreiber des Monitoringsystems versorgt EZS die Anlagenbetreiber, Techniker und andere Endbenutzer mit Alarmen und Berichten, die auf diese Zielgruppen zugeschnitten sind. Die Hauptverantwortung für die Umsetzung dieser Strategie liegt daher bei EZS. Die Berichte werden betriebliche und wirtschaftliche Daten enthalten und bilden damit ein wichtiges Instrument für die Beurteilung der Anlagenfunktion durch den Betreiber.

## 2 Vorgehen und Methode

### 2.1 Datenquellen für das Monitoring

Die Datenquelle ist entscheidend für fundierte Bewertungen von Prozessen und Systemen und die darauf basierenden Handlungsentscheidungen. Im *Abschnitt 2.1.1* werden die Möglichkeiten zur Installation des Temperatursensors dargestellt. In *Abschnitt 2.1.2* werden die für die Analyse erforderlichen Informationen aus den komplementären Wetterdaten aufgelistet.

#### 2.1.1 Instrumentierung von Anlagen mit dem Temperatursensor

*Abbildung 1* zeigt zwei Schemas von Solaranlagen mit 1) externem Plattenwärmeübertrager und 2) mit Glattrohrwärmeübertrager, der im Speicher integriert ist. Je nach Zugänglichkeit wird der IoT-Temperatursensors beim Kollektorfeld a) oder im Technikraum b) montiert. Der Temperatursensor verfügt über eine Batteriespeisung. Die Sendeleistung wird in gewissen Grenzen automatisch angepasst. Standardmässig ist ein Messintervall von 15 min eingestellt. Das Messintervall lässt sich jedoch mittels bidirektionaler Kommunikation verändern. Die Temperaturwerte werden über LoRaWAN in einer Datenbank gespeichert.

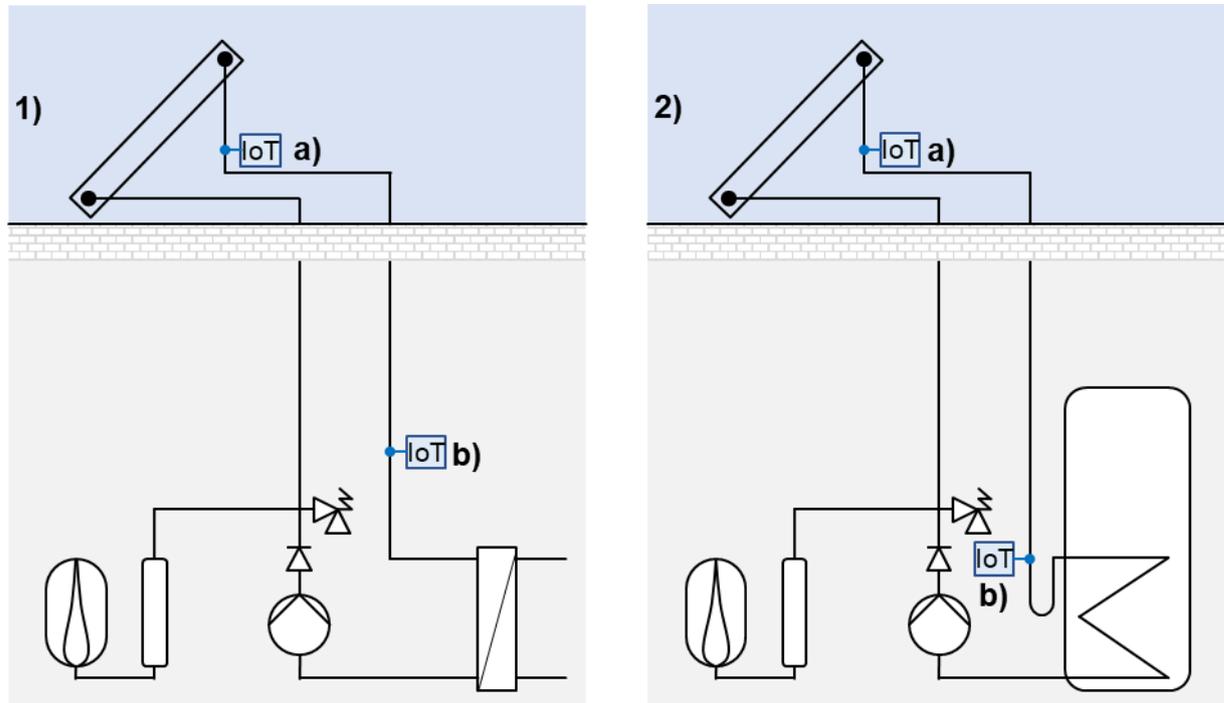


Abbildung 1 Solaranlage mit 1) externem Plattenwärmeübertrager und 2) mit Glattrohrwärmeübertrager im Speicher integriert. Die beiden möglichen Positionen des IoT-Temperatursensors sind mit a) und b) bezeichnet

Hinweis: Der Standort des Sensors und die Orientierung des Kollektors wurden nicht immer in dem für die Auswertung verfügbaren Datensatz erfasst.

### 2.1.2 Wetterdaten

Die Mehrzahl der Anlagen befindet sich in der Region Basel, daher wurden die Wetterdaten entweder von der Wetterstation Binningen-BL oder von der nächstgelegenen Station in einem 10 km-Netz verwendet. Diese Daten sind nicht frei verfügbar, aber im Monitoringdienst von EZS enthalten. Es wurde davon ausgegangen, dass aufgrund des dichten Netzes keine signifikanten Abweichungen entstehen, und es wurde keine Abweichungsanalyse zum Referenzstandort durchgeführt. Folgende Wetterdaten werden als Stundenwerte in dieselbe Datenbank geschrieben und den jeweiligen Temperaturwerten zugeordnet:

- Umgebungstemperatur
- Globale Bestrahlungsstärke auf die horizontale Ebene
- Bewölkungsgrad

## 2.2 Produktivsystem und Entwicklungssystem

EZS betreibt ein Produktivsystem, bestehend aus einer «Influx-Datenbank» (<https://www.influxdata.com/>) und dem RBA aus dem vorangegangenen Projekt, der in «NodeRed» (<https://nodered.org/>) entwickelt wurde. Zur Visualisierung der Daten wird «Grafana» (<https://grafana.com/>) verwendet. Die Datenbank enthält Zeitreihen von Sensordaten von jeder Anlage und die zugehörigen Wetterdaten, die über einen Online-Dienst von der nächstgelegenen Wetterstation gelesen werden. Ausserdem enthält die Datenbank Kundendaten und Ortsangaben. Diese Datenbank wächst mit jedem hinzugefügten Sensor. Für die Weiterentwicklung der Algorithmen wurde ein



Entwicklungssystem eingerichtet. Das Entwicklungssystem ist weitestgehend ein Klon des Produktivsystems, wobei die kundenspezifischen Angaben entfernt wurden.

Abbildung 2 zeigt schematisch die Struktur des Produktiv- und Entwicklungssystems sowie die Wirkungszusammenhänge zwischen den einzelnen Elementen. Abbildung 3 zeigt das Verfahren in Bezug auf die Datenbanken. Die Bedeutung der Wirkungszusammenhänge wird im folgenden Abschnitt erläutert.

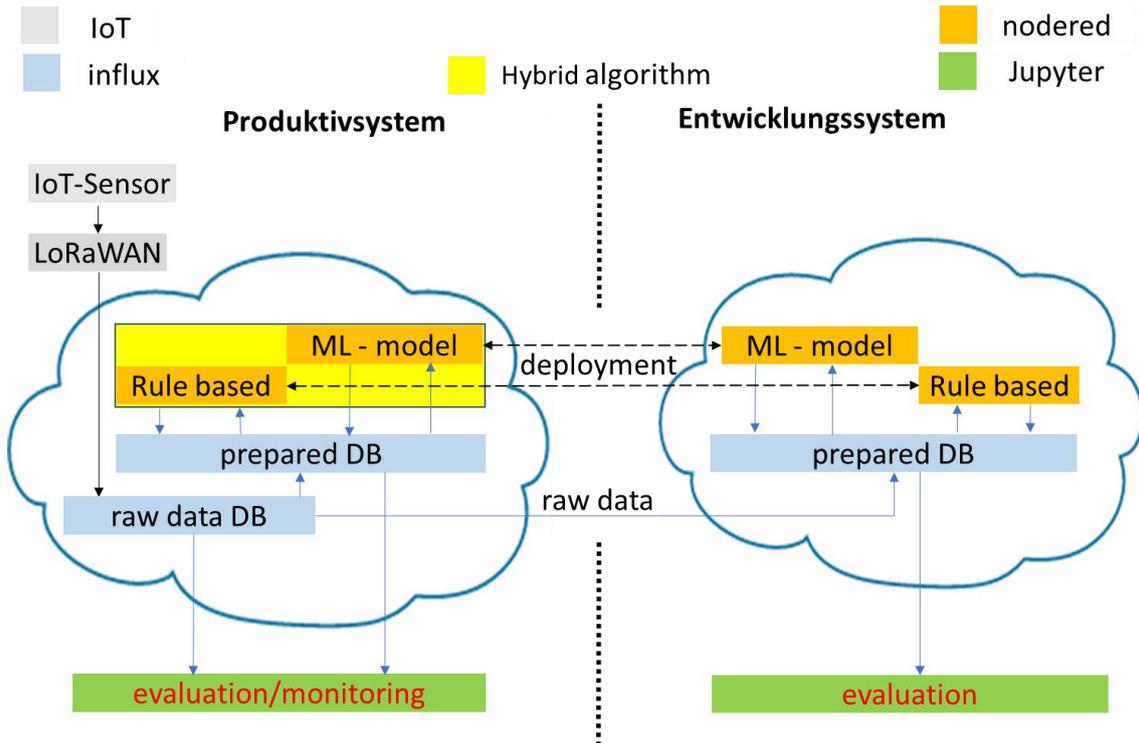


Abbildung 2 Die schematische Darstellung des symmetrischen Aufbaus von Produktivsystem und Entwicklungssystem

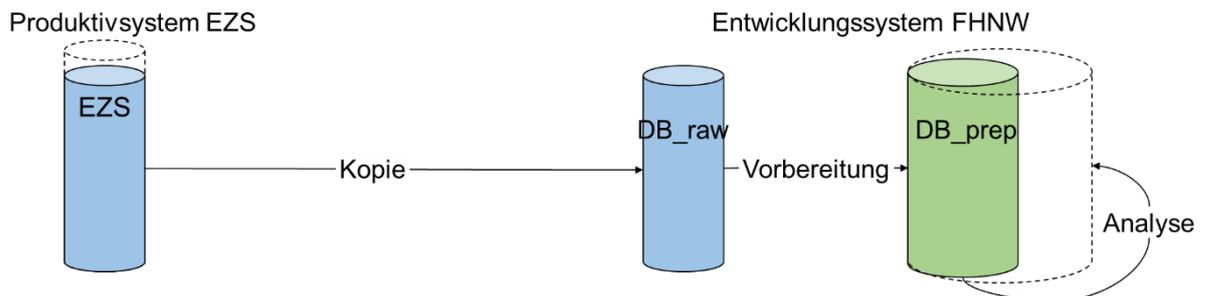


Abbildung 3 Datenbanken im Produktivsystem und im Entwicklungssystem

Basierend auf der Erfahrung im vorangegangenen Projekt «LoCoSol» wurde der bestehende Entwicklungsprozess wie folgt erweitert.

- Zwischen den Projektpartnern wurde ein optimierter Entwicklungsprozess vereinbart. Dieser regelte den Austausch von Daten und Softwareteilen, die Datensicherheit und die Versionskontrolle der Programme.



- Die gespiegelte Datenbank der Simulationsumgebung wurde regelmässig aktualisiert, sodass die Daten von neu in das Monitoring aufgenommenen Anlagen verfügbar waren. Die Daten in der Simulationsumgebung waren aus Datenschutzgründen anonymisiert. Die Zuordnung der Datensätze in der operativen Datenbank und in der Simulationsumgebung wurde über eine Nummer ermöglicht.
- Definition eines Prozesses/Interfaces für den Transfer des RBA zwischen der operativen Datenbank auf dem EZS-Server und der Simulationsumgebung, inklusive Versionskontrolle. Dadurch wurde erreicht, dass Programmfehler rasch behoben und Verbesserungen mit kleinem Aufwand implementiert werden konnten.
- Eine Testumgebung für den RBA wurde in «Python» (<https://www.python.org/>) entwickelt, wobei Standard-Datenanalysepakete wie «Pandas» (<https://pandas.pydata.org/>) verwendet wurden. Dies ermöglichte ein schnelles Testen des Algorithmus im Vergleich zu den technisch und administrativ aufwendigen Tests in der NodeRed-Grafana Umgebung.

Die Vorbereitung der Messdaten umfasst die Prozessschritte «Data Collection» und «Data Cleaning» (s. Abschnitt 3.1). Die Prozessschritte unter «Deployment» umfassen die Implementierung der Algorithmen auf dem Produktivsystem und die Kontrolle und Überwachung durch EZS im laufenden Betrieb (s. Abschnitt 3.7). Hier hat EZS die folgenden Möglichkeiten:

- Die Node-Red-Umgebung mit dem entwickelten und getesteten Algorithmus direkt zu nutzen
- Die logische Struktur des entwickelten und getesteten Algorithmus in die eigene Influx-Datenbankstruktur mit «Flux» (<https://www.influxdata.com/products/flux/>) zu implementieren

## 2.3 Datensätze für die Entwicklung der Algorithmen

Im vorangegangenen Projekt wurde sehr viel Erfahrung gesammelt über Datenmuster, die auf besondere Betriebszustände und Fehler schliessen lassen. Auf der Basis dieser Erfahrung wurde ein wesentlich leistungsfähigerer Prozess definiert, mit dem die Daten klassifiziert und für die Weiterentwicklung der Algorithmen aufbereitet werden. Dieser Prozess umfasst die folgenden Elemente:

- Die Datensätze von Anlagen, die bereits im vorangegangenen Projekt für das Training und die Validierung verwendet wurden, werden in anonymisierter, ansonsten unveränderter Form in einer separaten Datenbank gespeichert.
- Der Prozess der Vorbereitung und des Säuberens der Daten wird auf die neu anfallenden, wesentlich umfangreicheren Datenmengen angepasst und weitgehend automatisiert.
- Um die Funktionalität der KPIs zu überprüfen, werden die Datensätze einiger solarthermischer Anlagen untersucht, welche eine hohe Datendichte, also wenig Ausfälle der Messsysteme aufweisen. Dazu werden im Sommer, Winter und Frühling eines Jahres die Daten von je zwei Wochen auf ihre Schlüssigkeit überprüft.
- EZS stellte unterschiedliche Datensätze für verschiedene Sensoren zur Verfügung. Für alle Sensoren war ein Datensatz von ca. Oktober 2020 bis Oktober 2021 und für einige spezifische Sensoren ein Datensatz von Dezember 2021 bis April 2022 und Juni 2022 bis Juni 2023 verfügbar.
- EZS hat im Laufe des Projektes unterschiedliche Sensoren evaluiert und gab ein direktes Feedback zu den Sensoren, die für bestimmte KPI-Tests verwendet werden sollten, z. B. für die Nachtauskühlung oder die Durchmischung des Speichers

Schlussendlich fiel die Wahl auf die Sensoren der Baureihen *ST4042*, *ST4046*, *ST4048*, *ST4059*, *ST9249*, *ST9291*, *ST12091*, *ST12105*, *ST12108*, *ST12118*, und *ST12123*.



## 3 Entwicklung des verbesserten regelbasierten Algorithmus

Während der kritischen Überprüfung des alten RBA mit zusätzlichen Daten und der Verbesserung des Auswertungsprozesses wurde festgestellt, dass der RBA selbst eine höhere Genauigkeit für KPIs und Alarme erreichen kann, wenn er anhand von anlagenspezifischen Merkmalen und Felderfahrungen sorgfältig abgestimmt und in einem robusten Rahmen implementiert wird.

In diesem Abschnitt werden die wichtigsten Änderungen am Datenaufbereitungsprozess, am internen Auswertungsprozess und an der RBA im Vergleich zum Vorgängerprojekt beschrieben.

### 3.1 Datenvorbereitung

Datensätze können Lücken aufweisen die Temperaturwerte können auch mit Rauschen behaftet sein. Zudem ist das Messintervall nicht bei allen Anlagen gleich. Daher wurden die folgenden Schritte unternommen, um die Daten aufzubereiten. Die Ergebnisse sind in der *Abbildung 4* dargestellt.

- **Lückenhafte Daten:** Bei der Datenaufbereitung im vorgängigen Projekt hat sich gezeigt das Abtastwerte in der Zeitreihenmessung für die Temperatur fehlen, sogenannte Funkfehler. Im ersten Schritt werden daher die Lücken identifiziert und durch **Interpolation der benachbarten Messwerte** gefüllt. Die Ursache von Funkfehler sind «Fading» infolge Abfalls der Feldstärke oder veränderlicher Dämpfung und der als «Jamming» bezeichnete Datenausfall durch Störsignale. Die in Technikräumen installierten Sensoren sind am häufigsten von Funkfehlern betroffen.
- **Unterschiedliche Abtastrate:** Nicht jeder Sensor arbeitet mit derselben Abtastrate. Um alle Sensoren mit demselben Algorithmus zu analysieren, ist es jedoch erforderlich, dass die Zeitintervalle identisch sind. Daher werden die originalen Zeitreihen in neue Zeitreihen mit identischen **Zeitintervallen von 10 Minuten** transformiert, inklusive der zugehörigen Wetterdaten. Damit enthält jeder **Tagesdatensatz** Zeitreihen mit einer Länge von **144 Datenpunkten**. Die Ergebnisse aus dem vorherigen Schritt werden entsprechend normiert und in der Datenbank als Datensatz mit der Bezeichnung «UNIFORM» gespeichert.
- **Rauschen:** Das Signal der Kollektortemperatur hat einen Rauschanteil, der entfernt werden muss, bevor man die Ableitung bilden kann. Dies geschieht durch Bildung des gleitenden Durchschnitts (**Moving Average – MOV**). Da diese Art der Filterung immer einen Zeitversatz mit sich bringt, wird dieser in einem darauffolgenden Prozessschritt korrigiert, sodass die zeitliche Synchronität des Tagesgangs mit den Wetterdaten wieder hergestellt ist. *Abbildung 4* zeigt beispielhaft einen lückenhaften Tagesdatensatz, der mit den obengenannten Prozessen bearbeitet wurde. Dieser bereinigte Datensatz wird zusammen mit den auf dasselbe Intervall transformierten Wetterdaten in eine separate Datenbank «prepared\_DB» geschrieben. Für die Berechnung des MOV wurde ein **zweistündiges Zeitfenster** implementiert.

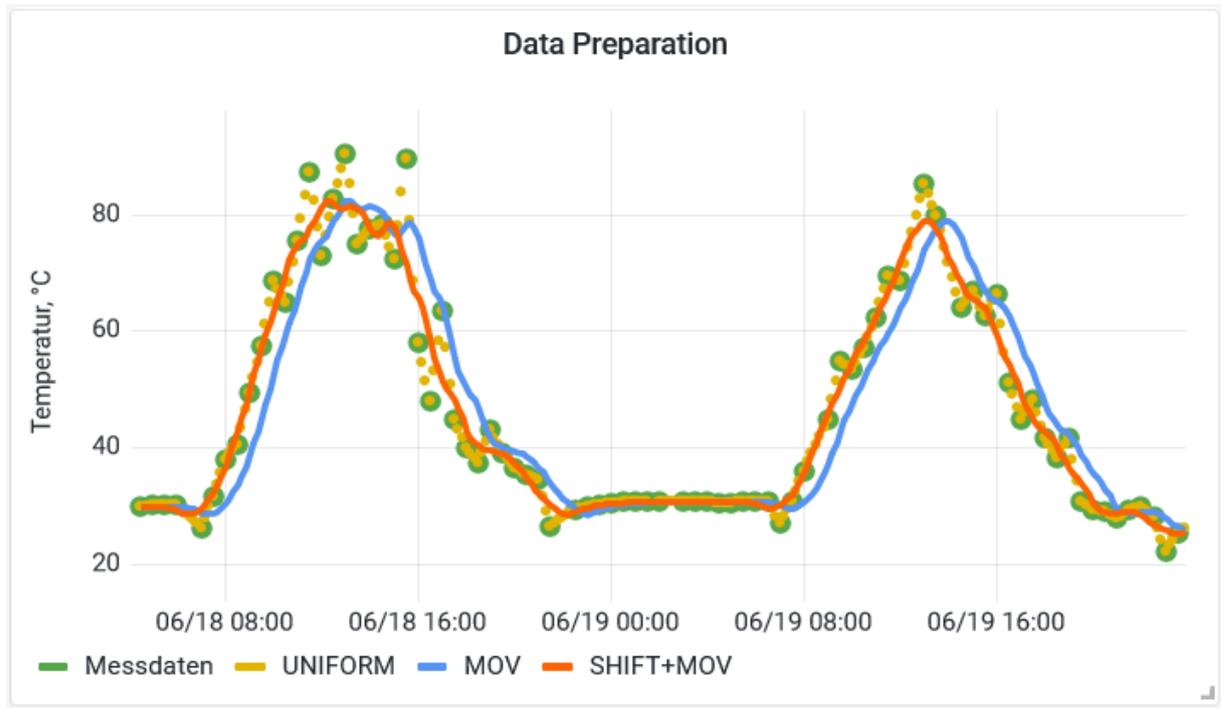


Abbildung 4 Lückenhafter Tagesdatensatz (Messdaten), uniformierten Daten mit 10 min. Intervall (UNIFORM), geglättete Daten (MOV), zurückverschobene Daten (SHIFT+MOV)

Diese Datenbank wird vom regelbasierten Algorithmus (und später auch vom MLA) gelesen, analysiert und mit den Ergebnissen beschrieben. Für dieses «pre-processing» wurde ein Softwaremodul in Node-RED entwickelt. Der Prozess wird täglich nach Mitternacht automatisch gestartet und die Daten werden für Auswertung in Datenbank gespeichert (s. *Abbildung 3*)

### 3.2 KPI-Definition

Die KPIs wurden entwickelt, um bestimmte Ereignisse während des Betriebs einer Solaranlage zu identifizieren, um die Leistung, den Wirkungsgrad und technische Probleme wie Überhitzung, Stagnation oder Nachtauskühlung des Speichers zu analysieren. Das bedeutet, dass der KPI ein messbares Ereignis beschreibt, das durch numerische Werte dargestellt werden kann. Das zugehörige Label bezeichnet den Datensatz, bei dem ein KPI auftritt. Diese werden durch den KPI-Generator des RBA generiert. Die KPIs sind so strukturiert, dass der Analyseteil des RBA-Fehlfunktionen erkennen und Aussagen zum Betriebszustand ableiten kann.

Einige KPIs wurden in dem vorherigen Projekt mithilfe eines logischen Algorithmus erstellt, welcher die Überwachungsdaten der Vorlauftemperatur verwendete. Für die Beschreibung der Temperaturverläufe werden folgende Vereinbarungen getroffen:

- Ein monotoner Verlauf innerhalb eines Intervalls ist dadurch gekennzeichnet, dass sich das Vorzeichen der zeitlichen Ableitung nicht ändert.
- Die Begriffe «Konvex» und «Konkav» beschreiben den monoton steigenden oder fallenden Temperaturverlauf im Vergleich zu einem hypothetischen linearen Verlauf mit denselben Anfangs- und Endtemperaturen.
- Bei einem monoton konkaven Anstieg liegen die Temperaturen innerhalb eines Intervalls oberhalb des hypothetischen linearen Anstiegs. Das bedeutet, dass die zeitlichen Ableitung des Temperaturverlaufs mit der Zeit kleiner wird.



- Bei einem monoton konvexen Anstieg liegen die Temperaturen innerhalb eines Intervalls unterhalb des hypothetischen linearen Anstiegs. Das bedeutet, dass die zeitliche Ableitung des Temperaturverlaufs mit der Zeit zunimmt.

Im Rahmen dieses Projekts wurden die Regeln und Parameter des vorherigen Algorithmus ausgiebig getestet. Die gesamte Programmierstruktur wurde deutlich verschlankt (Redundanzfrei, deskriptive Variablen- und Funktionsnamen, Benutzerfreundlich) und einige zusätzliche KPIs wurden hinzugefügt. Die **wichtigsten Neuerungen des RBA** im Vergleich zur vorherigen Version sind:

- Verbesserte Filterung der täglichen Datenmenge unter Berücksichtigung der Wetterbedingungen und der Tageszeit in Kombination mit der Temperatur des Solarkollektors,
- Optimierte Anpassung der Parameter wie Zeitperioden und logischen Schwellenwerte mithilfe visueller und statistischer Analysen vergangener Daten,
- Bessere Interpretation der Temperaturableitung,
- Verbesserte Softwareumgebung für robustere und schnellere Tests und Anwendungen.

Im folgenden Abschnitt werden einige Beispiele für die RBA-Implementierung angeführt. Sie entstammen dem «Pflichtenheft»<sup>2</sup>.

Folgende **Herausforderungen** waren bei der Definition der KPIs zu beachten:

- Begrenzte Verfügbarkeit von zuverlässigen Daten,
- Anstrengende Tests in einer visuellen Umgebung,
- Interpretation der verschiedenen physikalischen Phänomene, die sich bei unterschiedlichen Ausführungen von solarthermischen Anlagen zeigen,
- Einschränkung der Verwendung von nur einem Datensensor pro Anlage,
- mehrere physikalische Phänomene innerhalb desselben KPI.

Basierend auf den aktuellen Projektanforderungen und den Ergebnissen früherer Projekt wurden insgesamt sieben KPIs definiert. Diese KPIs sind in *Tabelle 1* unten aufgeführt:

Tabelle 1 Tabellarische Übersicht der KPIs

KPI-ID	Name (Beschreibung unten im Text)	Ausgabe
<b>Anlagebetrieb</b>		
N1	Erster Anstieg	Zeitstempel
N2	Pumpenstart	Zeitstempel
N3	Rasche Steigungsänderung	Zeitstempel
N4	Maximaltemperatur am Tag	Zeitstempel und entsprechende Temperatur
<b>Nachtauskühlung</b>		
N5	Normalbetrieb	Ja oder Nein Minimaltemperatur an der Nacht
N6	Nachtauskühlung (Schwerkraftzirkulation oder Umwälzpumpe)	Ja oder Nein
N7	Rohrinterne Nachtauskühlung	Ja oder Nein

<sup>2</sup> Weitere technische Details zu diesem Dokument sind bei EZS erhältlich.



Die KPIs N1 bis N4 sind relevant für die Ermittlung der Start-Stopp-Phasen der Anlage und der wichtigsten Betriebsparameter.

1. N1: Der KPI-N1 dient der Identifizierung des ersten signifikanten Temperaturanstiegs des Kollektors am Morgen. Er wird ausgegeben, sobald die Differenz einer Vorlauftemperatur zur vorhergegangenen Temperaturmessung mehr als der vordefinierte Grenzwert beträgt.
2. N2: Der KPI-N2 markiert den Zeitpunkt, bei welchem die Pumpe der Anlage gestartet wird, und ist relevant für den weiteren Anlagenbetrieb des Tages.
3. N3: Der KPI-N3 markiert den Zeitpunkt, an welchem der Pumpenbetrieb gestoppt wird.
4. N4: Der KPI-N4 markiert den Zeitpunkt des Tages, an welchem die Kollektortemperatur ihren Höchststand erreicht.

Die KPIs N5 bis N7 sind relevant für die Erkennung von Zuständen, die auf unerwünschte Zirkulation hindeuten. Diese kann infolge Dichteunterschiede durch Schwerkraft getrieben sein. Die Ursache ist meist ein undichtes Rückschlagventil. Dadurch entstehen teilweise erhebliche Wärmeverluste und verringern somit den Gesamtwirkungsgrad der Anlage. Diese Phänomene werden unter dem Begriff «Nachtauskühlung» zusammengefasst.

5. N5: Der KPI-N5 sucht nicht die Nachtauskühlung selbst, sondern prüft umgekehrt, ob es sich um einen normalen Verlauf der Vorlauftemperatur handelt. Durch die Analyse dieser Information wird festgestellt, ob sich der Fühler am Kollektoranschluss oder an der Vorlaufleitung im Technikraum befindet. Anschliessend wird die Differenz zwischen der Sensortemperatur und der Umgebungstemperatur ermittelt. Bei korrekter Funktion der Anlage und Installation des Sensors im Technikraum kann die Minimaltemperatur des Sensors als Temperatur des Technikraums interpretiert werden.
6. N6: Der KPI-N6 dient zur Identifikation der Nachtauskühlung aufgrund von fehlerhaften Hardwarekomponenten wie Rückschlagventil, Temperatursensor oder Regler ermitteln. Nachtauskühlung wegen undichtem Rückschlagventil führt zu Schwerkraftzirkulation. Ein defekter Temperatursensor oder defekter Regler kann einen Dauerbetrieb der Pumpe zur Folge haben, was ebenfalls zur Nachtauskühlung aufführt.
7. N7: Der KPI-N7 dient zur Identifikation von Wärmeverlusten durch rohrinterne Zirkulation, die nur Rohrleitungen im Technikraum betreffen, nicht aber das Kollektorfeld.

Die KPI1 bis KPI4 wurden aus dem vorangegangenen Projekt übernommen. Zur Unterscheidung zwischen neuen und alten KPIs wurde die unterschiedliche Nomenklatur beibehalten.

- KPI1: Kein LoRaWAN Signal
- KPI2: Batterielevel tief
- KPI3: Temperatur > 100 °C
- KPI4: Temperatur > 100 °C an aufeinanderfolgenden Tagen

Um die Entwicklung des Codes zu vereinfachen, wurden technische Voraussetzungen für die Datenanalyse festgelegt. Diese basieren auf der Betriebserfahrung und dem Fachwissen über die verschiedenen physikalische Phänomene, die in einer solarthermischen Anlage auftreten. Dazu gehört insbesondere die Erfahrung, die mit dem Einsatz von IoT Temperatursensoren gemacht wurden. Es hat sich bereits früh herausgestellt, dass die Kollektorfelder nur in Ausnahmefällen leicht für die Sensormontage zugänglich sind. Seither werden alle Sensoren standardmässig im Technikraum installiert. Die Datensätze sind entsprechend gekennzeichnet. Die Voraussetzungen für die Datenanalyse lauten:



- Die KPI-Analyse für N1...N4 für den Stichtag beginnt um 07:00 Uhr des jeweiligen Tages,
- Die KPI-Analyse für N5...N7 für den Stichtag beginnt um 19:00 Uhr des jeweiligen Tages bis 06:00 Uhr des nächsten Tages,
- Es werden nur Datensätze berücksichtigt, bei denen sich der Temperatursensor im Technikraum befindet,
- Da die Orientierung des Kollektors nicht in allen Datensätzen verfügbar war, wurde sie nicht in die KPI-Logik implementiert.

Für jeden KPI stellte EZS dem FHNW-Team einige Beispiele aus der Produktivsystem zur Verfügung, die das zu identifizierende Verhalten markieren. Weitere typische Muster wurden in regelmässigen Sitzungen besprochen und dokumentiert.

Die allgemeine Logik zur Ermittlung dieser KPIs wird in den folgenden Unterabschnitten beschrieben.

### 3.2.1 Logik für N1 (ersten signifikanten Temperaturanstiegs)

Wie in *Abbildung 5* zu sehen ist, wurden vier verschiedene Szenarien vorgegeben, die mit «N1» gekennzeichnet werden sollen. Wenn die Temperaturdifferenz zwischen zwei Zeitschritten zum ersten Mal am Tag einen Schwellenwert (Einstellbarparameter) überschreitet, wird «N1» ermittelt.

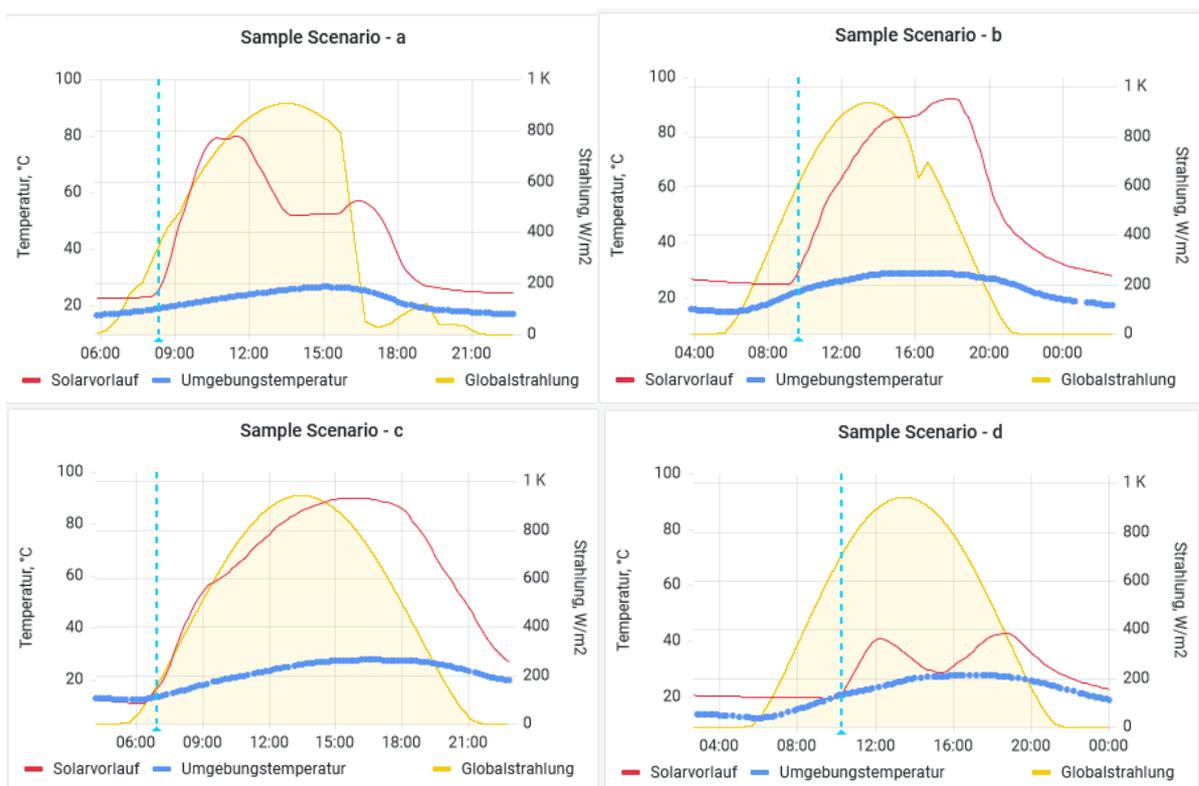


Abbildung 5 Beispiele für verschiedene Szenarien, in denen N1 beim ersten Anstieg identifiziert werden sollte

### 3.2.2 Logik für N2 (Pumpenstart)

Wie in *Abbildung 6* dargestellt, wurden ebenfalls vier verschiedene Szenarien definiert, die mit «N2» gekennzeichnet werden sollen. Diese stellen einen typischen Betriebsbereich (monoton konkaven Anstieg) nachdem die Pumpe startet, dar.



- Szenario-a und Szenario-b: Ein monotoner, konkaver Temperaturanstieg über einen bestimmten Zeitraum, bis eine rasche Temperaturänderung beobachtet wird. Ein bestimmter Punkt am Anfang dieser Kurve kann als Pumpenstart festgelegt werden.
- Szenario-c: Ein vollständig monotoner, konkaver Temperaturverlauf bis zum Sonnenuntergang oder einer geringeren Sonneneinstrahlung. Ein bestimmter Punkt am Anfang dieser Kurve kann als Pump-Start festgelegt werden.
- Szenario-d: Ein Sonderfall, in dem keiner der oben genannten Fälle eintritt, obwohl ein anfänglicher Temperaturanstieg zu verzeichnen ist. Der Sonderfall ist mit dem gleichen KPI zu identifizieren.

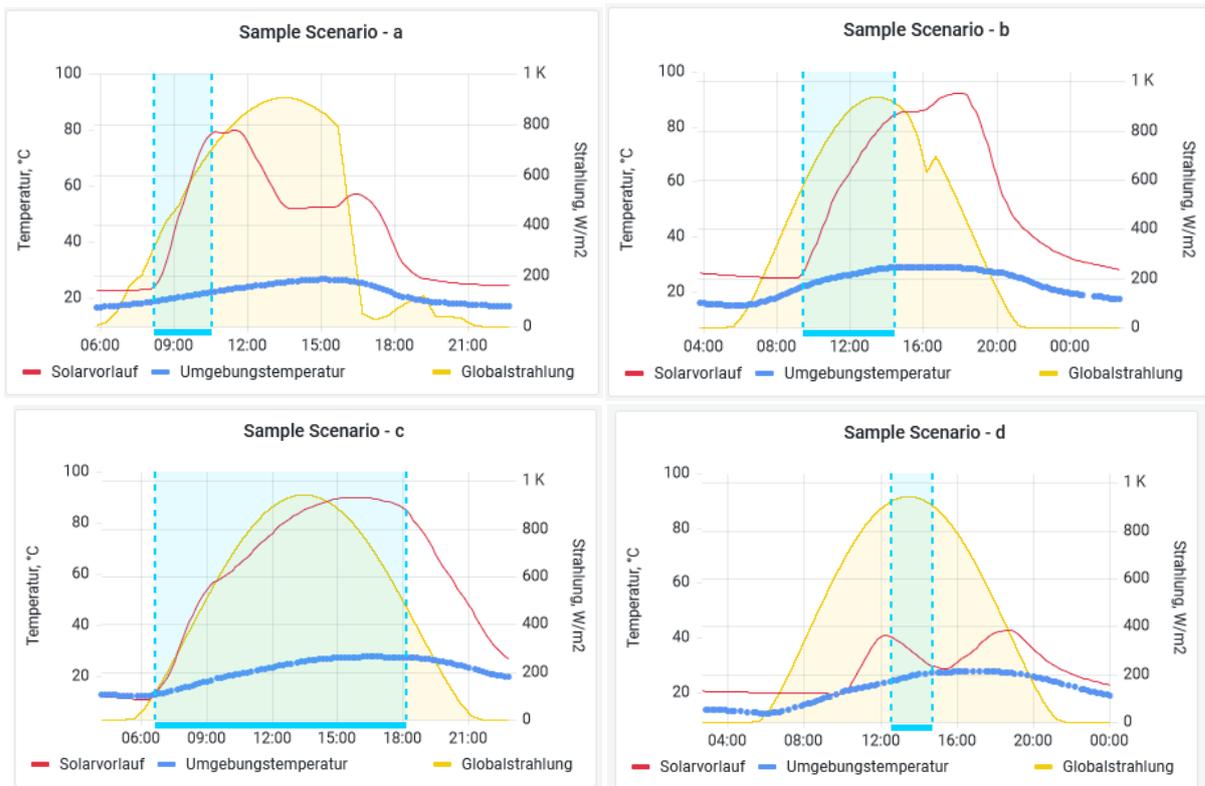


Abbildung 6 Beispiele für verschiedene Szenarien, in denen N2 als Pumpenstart im Verhältnis zu N1 identifiziert werden sollte. Der blau schattierte Bereich in den Szenarien a, b, und c zeigt den erwarteten Betriebsbereich, in dem die Kollektortemperatur monoton konkav ansteigt. Szenario-d ist Sonderfall

Die folgende logische Struktur (Abbildung 7) wurde implementiert für Identifikation des Pumpenstarts anhand eines stetigen Anstiegs der Temperatur mit konkavem Verlauf über eine gewisse Zeitspanne. Abrupte Änderung der Steigung in positiver oder negativer Richtung wurde auch berücksichtigt.

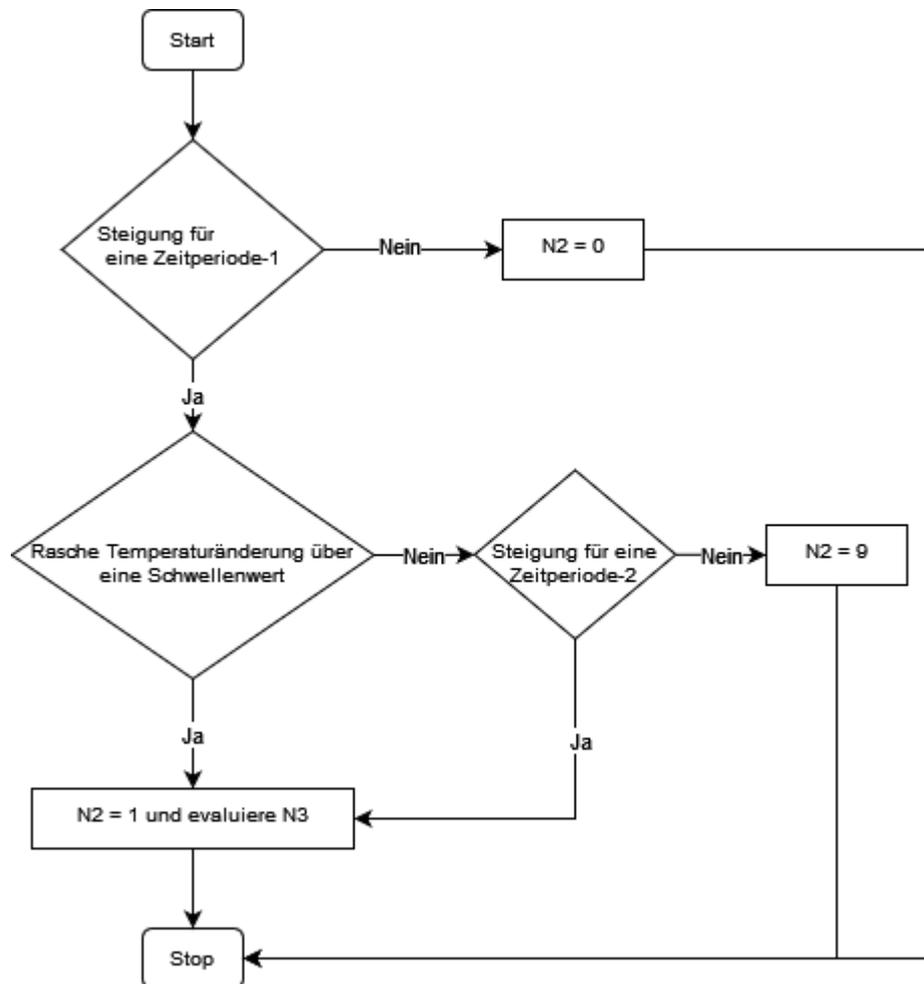


Abbildung 7 Logische Struktur für die Erstellung von KPI «N2»

### 3.2.3 Logik für N3 (Pumpenstopp)

Wie in *Abbildung 8* dargestellt wurden vier verschiedene Szenarien definiert, die mit «N3» gekennzeichnet werden sollen. Diese stellen ein typisches Ende oder eine Unterbrechung des Betriebs einer Solaranlage dar.

- Szenario-a und Szenario-b: Eine rasche Temperaturänderung nach einer bestimmten Betriebszeit. An diesem Punkt kann das Betriebsende festgelegt werden.
- Szenario-c: Eine vollständig monotone, konkave Temperaturkurve bis zum Sonnenuntergang oder einer geringeren Sonneneinstrahlung. Ein bestimmter Punkt am Ende dieser Kurve kann als Betriebsende festgelegt werden.
- Szenario-d: Ein Sonderfall, bei dem die solare Vorlauftemperatur nach einer bestimmten Betriebszeit 100 °C überschreitet. Dies wird als Betriebsende bezeichnet.

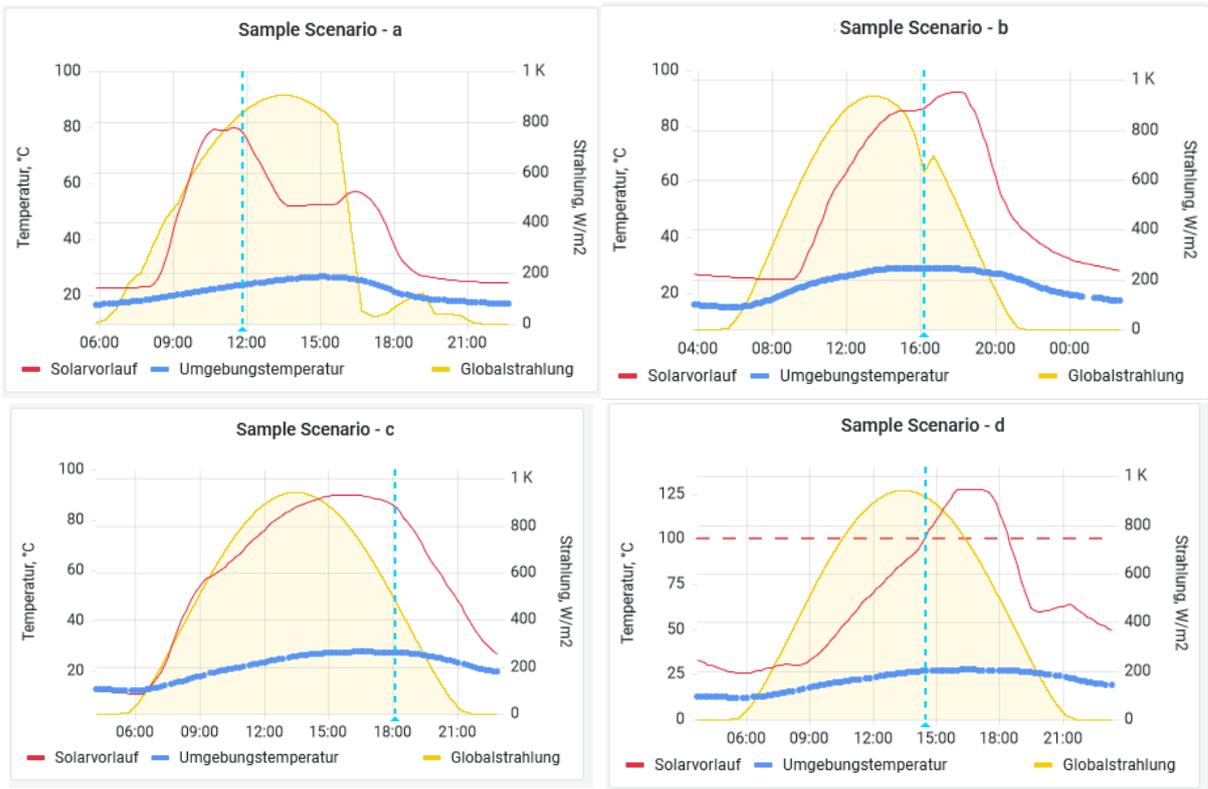


Abbildung 8 Beispiele für verschiedene Szenarien, in denen verschiedene Arten von N3 identifiziert werden sollten

Die folgende logische Struktur (Abbildung 9) wurde implementiert.

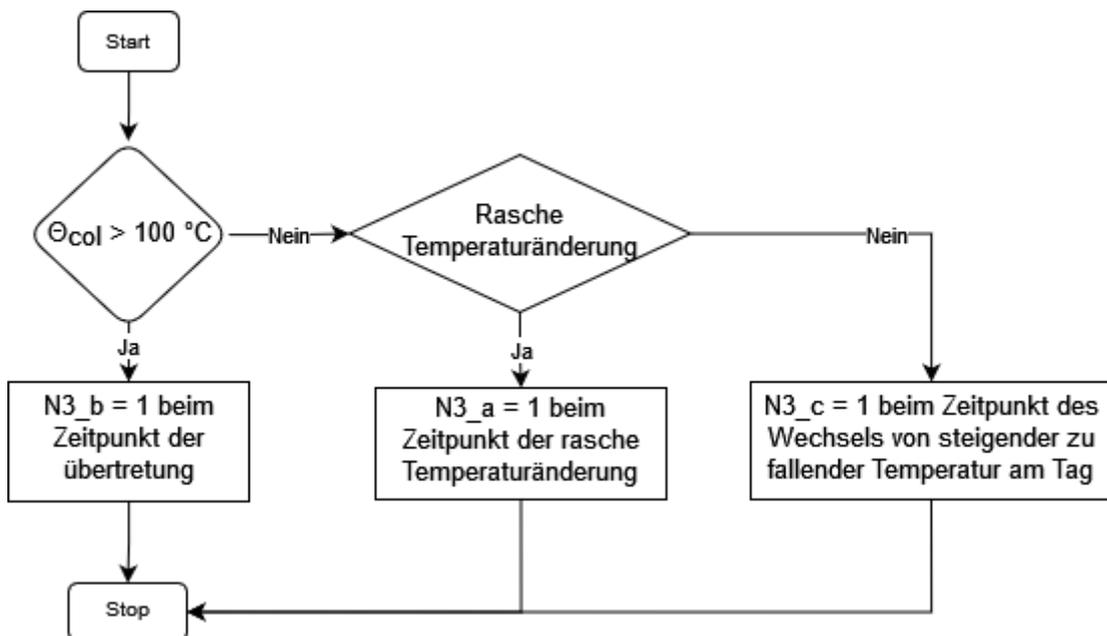


Abbildung 9 Logische Struktur für die Erstellung von KPI «N3»



### 3.2.4 Logik für N4 (max. Vorlauftemperatur)

«N4» ist die maximale gemessene Temperatur am Tag. Wie in *Abbildung 10* unten zu sehen ist, wird dieser Wert unabhängig vom Verlauf der Temperatur des Sonnenkollektors gespeichert. An einem guten Betriebstag jedoch, wie in der ersten Abbildung gezeigt, wird die erreichte Maximaltemperatur ausgewertet, um eine Abschätzung der Speichertemperatur nach einer Tagesladung zu erhalten. Diese kann Aufschluss darüber geben, ob das Speichervolumen zu 100% oder nur teilweise genutzt wurde.

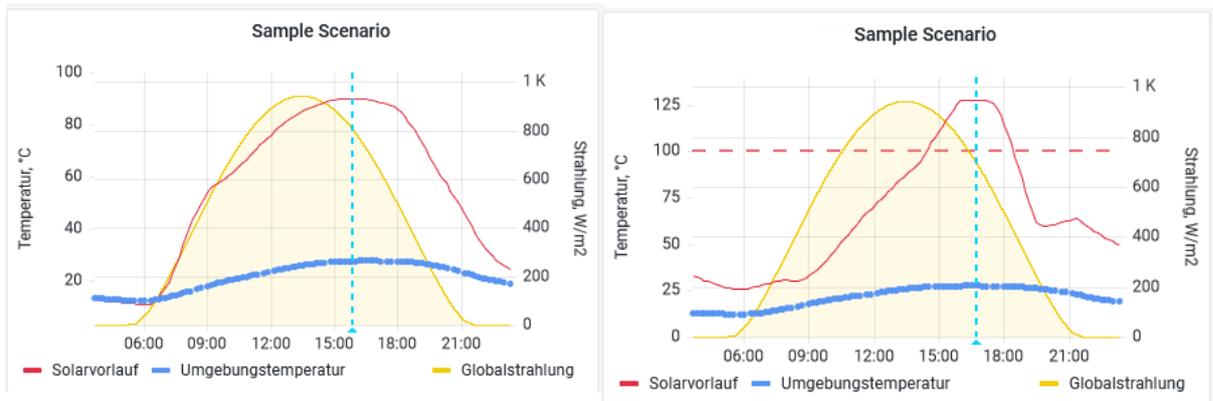


Abbildung 10 Beispiele für die Festlegung von N4 unabhängig von einem Muster

### 3.2.5 Logik für N5 (Betrieb ohne Nachtauskühlung)

N5 steht für den Betrieb ohne Nachtauskühlung und die Anlage funktioniert einwandfrei. Die Auswertung erfolgt von 22.00 Uhr am Stichtag bis 07.00 Uhr am Folgetag, umfasst also eine Nacht. In diesem Auswertebereich wird die Gesamtabweichung der Kollektortemperatur überprüft. Ist diese Abweichung kleiner als ein Schwellenwert und liegt kein N6 oder N7 vor, dann ist N5 = 1, ansonsten ist N5 = 0.

Wenn N5 an drei aufeinander folgenden Tagen aktiviert ist, kann, wie in der *Abbildung 11* unten gezeigt, der Mittelwert der minimalen Temperatur dieser Tage als Kellertemperatur geschätzt werden. Diese Logik wird direkt im Produktivsystem implementiert.

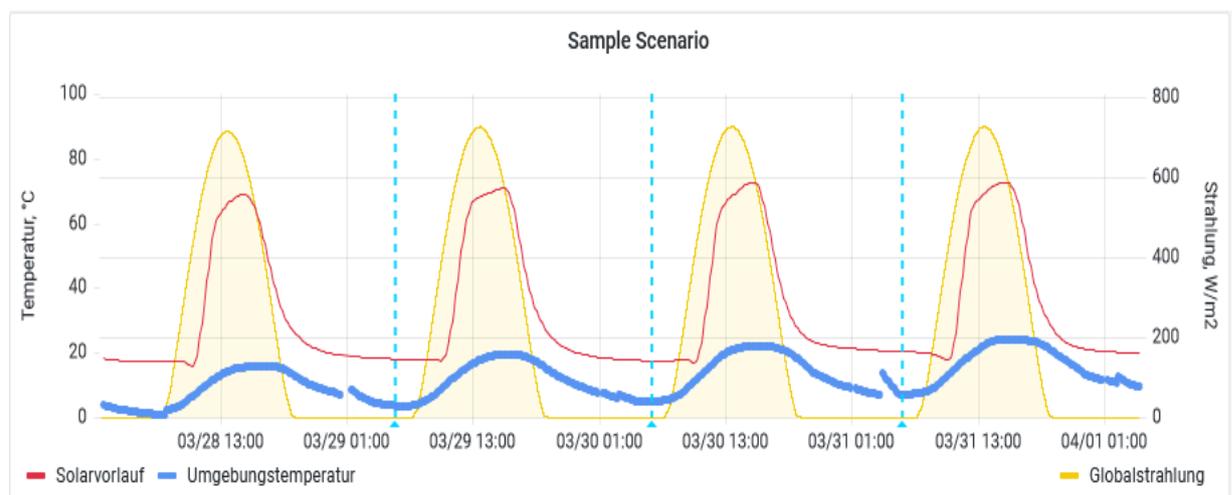


Abbildung 11 Beispiel für die Festlegung von N5 zum Zeitpunkt der Mindesttemperatur in der Nacht, wenn keine Nachtauskühlung festgestellt wurde. Die Werte von drei aufeinanderfolgenden Nächten können zur Schätzung der Kellertemperatur verwendet werden



### 3.2.6 Logik für N6 (Nachtauskühlung nach Aussen)

Mit N6 werden die folgenden zwei Arten der Nachtauskühlung erfasst.

- **Undichtiges Rückschlagventil:** Nach dem Pumpenstopp ist die Vorlaufleitung wärmer als die Rücklaufleitung. Aufgrund des Dichteunterschiedes wird die Flüssigkeit entgegen der Pumprichtung beschleunigt. Warme Flüssigkeit tritt aus dem vorlaufseitigen Anschluss des Wärmeübertragers, fließt zum Kollektor, wird abgekühlt, durchströmt die ausgeschaltete Pumpe und tritt in den unteren Anschluss des Wärmeübertragers. Der Fühler misst daher eher hohe Temperaturen, beispielsweise 30...50 °C die langsam abnehmen. Ein Siphon allein kann die Nachtauskühlung nicht verhindern. Dieser Fall ist in der *Abbildung 12* dargestellt und durch «N6\_a» definiert.
- **Ungewollter Pumpenbetrieb:** Die Flüssigkeit kommt aus dem Wärmeübertrager des Speichers. Ihre Temperatur liegt im Bereich der unteren Speichertemperatur und ist daher eher tief, beispielsweise 30...40 °C. Die Flüssigkeit durchströmt den Kollektor, kühlt sich ab und tritt über die Vorlaufleitung wieder in den Wärmeübertrager. Wenn keine Zusatzheizung vorhanden ist oder der Speicher nicht geladen ist, liegen die gemessenen Temperaturen nur geringfügig über der Umgebungstemperatur, beispielsweise 5...10 K. Bei aktivierter Zusatzheizung oder einem vollgeladenen Speicher kommt es zur Durchmischung, und die gemessenen Temperaturen liegen dauerhaft deutlich über der Umgebungstemperatur, beispielsweise 10...20 K. Dieser Fall ist in der *Abbildung 13* dargestellt und durch «N6\_b» definiert.

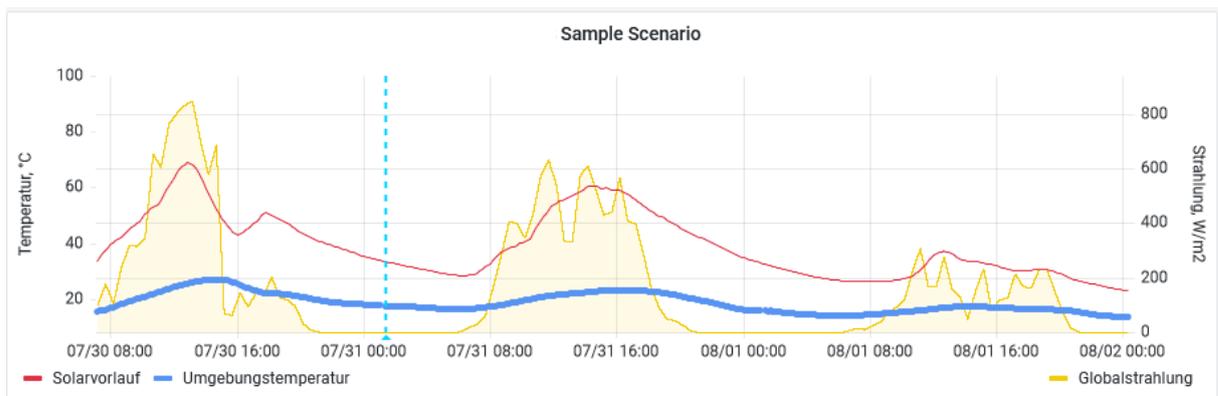


Abbildung 12 Beispiel für eine Nachtauskühlung aufgrund eines defekten Rückschlagventils (N6\_a)

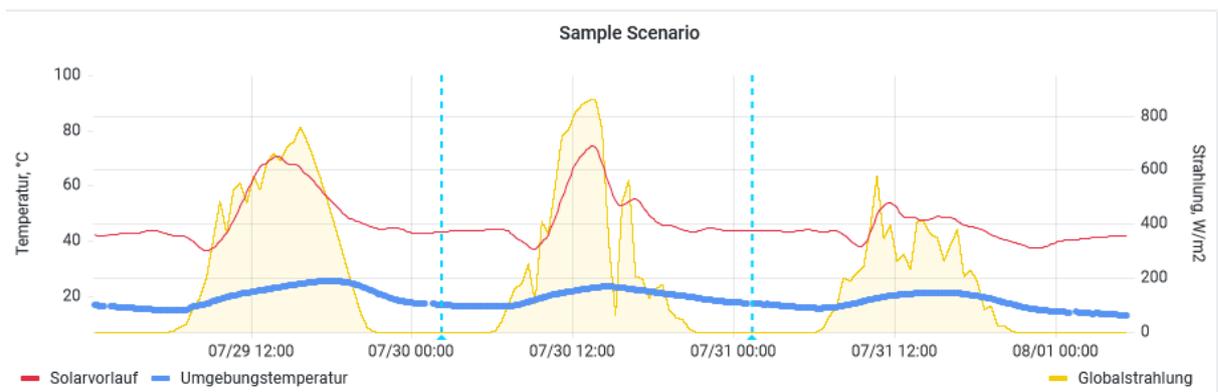


Abbildung 13 Beispiel für eine Nachtauskühlung aufgrund eines defekten Pumpenregelung beziehungsweise ungewollter Pumpenbetrieb (N6\_b)



Die folgende Logik ist implementiert, um N6\_a und N6\_b zu identifizieren.

- N6\_a: Wenn die Temperaturdifferenz zwischen Vorlauf- und Aussentemperatur über den gesamten (nächtlichen) Auswertebereich grösser als ein Schwellenwert bleibt und die Standardabweichung der Vorlauftemperatur grösser als ein Schwellenwert ist, dann ist N6\_a = 1, sonst ist N6\_a = 0. Die Schwellenwerte sind einstellbare Parameter und anstelle der Aussentemperatur könnte auch die Kellertemperatur verwendet werden. Deshalb wird sie im Produktivsystem aus dem Mittelwert von drei aufeinanderfolgenden N5-Tagen bestimmt.
- N6\_b: Wenn die Vorlauftemperatur während des Auswertungsbereichs **immer** über einem maximalen Schwellenwert liegt und die Standardabweichung der Vorlauftemperatur kleiner als ein Schwellenwert ist, dann ist N6\_b = 1, sonst N6\_b = 0

### 3.2.7 Logik für N7 (Rohrinterne Zirkulation)

Durch **Rohrinterne Zirkulation** wird Wärme aus dem Speicher zum Temperatursensor transportiert. Verglichen mit anderen Arten der Nachtauskühlung (N6\_a und N6\_b) ist der Wärmeverlust des Speichers klein, sodass auch die Sensortemperatur nur langsam abnimmt. *Abbildung 14* zeigt eine Nachtauskühlung während einer Schönwetterperiode. Bei anschliessendem Schlechtwetter (hier nicht gezeigt) bleibt die Vorlauftemperatur konstant.

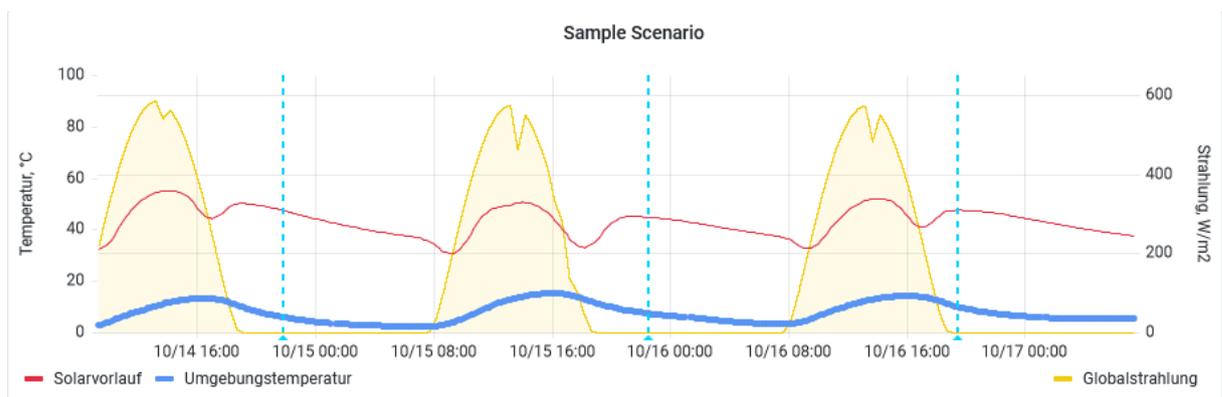


Abbildung 14 Beispiel für eine Nachtauskühlung aufgrund eines Rohrinterne Zirkulation

Für die Identifizierung von N7 wird die gleiche Logik wie bei N1 und N2 angewandt, allerdings nur für den Nachtbereich. Wenn es also nach 19:00 Uhr zu einem Temperaturanstieg kommt, der mindestens für einen vordefinierten Zeitraum anhält, dann ist N7 = 1, sonst N7 = 0.

Um so viele Informationen wie möglich über die Ursache der Nachtauskühlung zu erhalten, wurden N6 und N7 nicht in einer logischen Struktur kombiniert, so dass beide gleichzeitig aktiviert werden konnten, wenn ihre jeweiligen Bedingungen erfüllt waren. Dies gibt die Möglichkeit, den tatsächlichen Grund für die Nachtauskühlung zu ermitteln.

## 3.3 Explorative Datenanalyse zur Parameterabschätzung

Die Variablen «Zeitperiode» oder «Schwellenwert» in den obigen Abschnitten sind Beispiele für einstellbare Parameter, die die Genauigkeit des Codes beeinflussen und anhand der visuellen Analyse der Ergebnisse und der statistischen Analyse der Messdaten abgestimmt wurden. Die statistische Analyse wurde auch für die ausgewählten Datensätze durchgeführt (siehe *Abschnitt 2.3*), um die folgenden Punkte zu bewerten:



- Prozentsatz der fehlenden Daten
- Ausschluss von Daten in Bezug auf den Parameter Wolkenbedeckung, Variation (2%, 10%, 20%)
- Veränderung der Anzahl der Alarme in Abhängigkeit vom Bewölkungsparameter, Variation (2%, 10%, 20%)
- Korrelation zwischen Vorlauftemperatur, Sonneneinstrahlung und Bewölkungsgrad.
- Identifizierung von Ausreißern
- Vergleich der Anzahl Alarme, die gemäss vorherigen Logik und dem optimierten RBA generiert wurden

So wurde beispielsweise der Wert für einen Schwellenwert für die zeitliche Ableitung der Temperatur zur Ermittlung des KPI-N1 auf der Grundlage eines Boxplot und Histogramms (siehe *Abbildung 15* und *Abbildung 16*) für den KPI-N1 auf 0.2 K/min geschätzt.

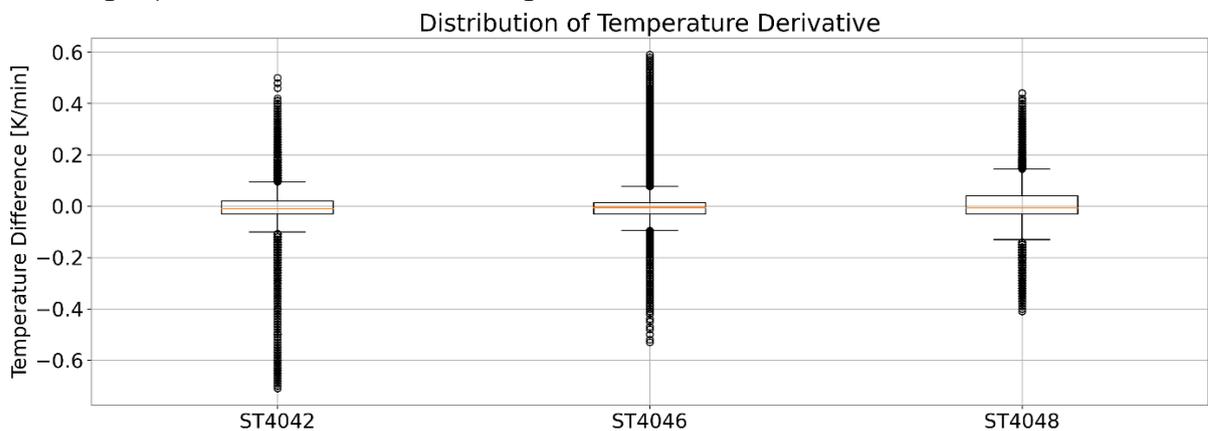


Abbildung 15 Beispiel für eine visuelle statistische Analyse zur Parameterschätzung

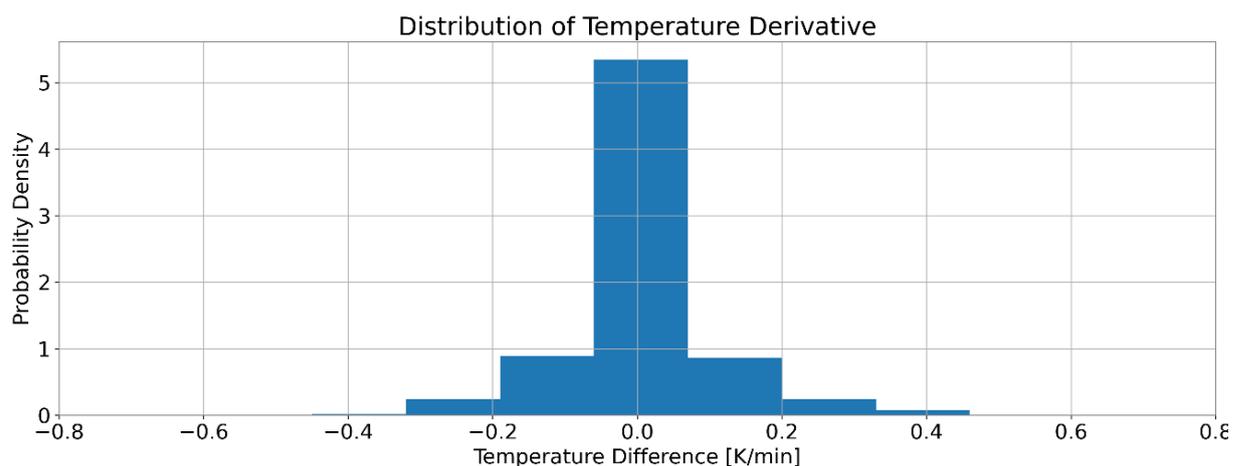


Abbildung 16 Beispiel für eine visuelle statistische Analyse zur Parameterschätzung

Zusätzlich wurde ein studentisches Projekt<sup>3</sup> zur elementaren Parameteranalyse mittels eines iterativen Verfahrens durch Veränderung einzelner Parameter und deren Kombinationen durchgeführt. Die

<sup>3</sup> Eine Kopie der Studienarbeit ist auf Anfrage bei der FHNW erhältlich



Ergebnisse dieser Arbeit dokumentieren ausführlich die Auswirkung der Wahl der Parameter auf die Genauigkeit der RBA und empfehlen einen bestmöglichen Parametersatz für die gegebenen Datensätze. Diese wurden in der endgültigen Version des Codes implementiert. Die iterative Methode zur Parameterschätzung ist ebenfalls dokumentiert und kann zur Verbesserung der RBA-Genauigkeit in ihrem Produktivsystem verwendet werden.

Es wurde auch die Möglichkeit geprüft, die zweite Ableitung der Temperatur zur Identifizierung signifikanter Ereignisse einzusetzen. Die Analyse hat jedoch gezeigt, dass die Ergebnisse der zweiten Ableitung ein verrauschtes Signal ergeben und die Grösse des Signals bei den verschiedenen Sensoren stark variiert. Daher war bei der zweiten Ableitung die Festlegung eines einzigen Schwellenwerts für alle Sensoren unpraktisch und führte zu Ungenauigkeiten.

In der *Abbildung 17* zeigt zum Beispiel die obere Kurve die gemessene Temperatur für den jeweiligen Sensor und die untere Kurve seine zweite Ableitung. Hier ist zu erkennen, dass für den Sensor *ST4042* ein Wert von «0.2 (K/min)/min» in der zweiten Ableitung abrupte Änderungen darstellt, während dieser Wert für den Sensor *ST4046* im Bereich des normalen Betriebs liegt. Diese sind durch die senkrechten roten Linien hervorgehoben. In ähnlicher Weise wird beim Sensor *ST4048* eine rasche Temperaturänderung bereits bei Grössenordnungen deutlich unter «0.2 (K/min)/min» festgestellt.

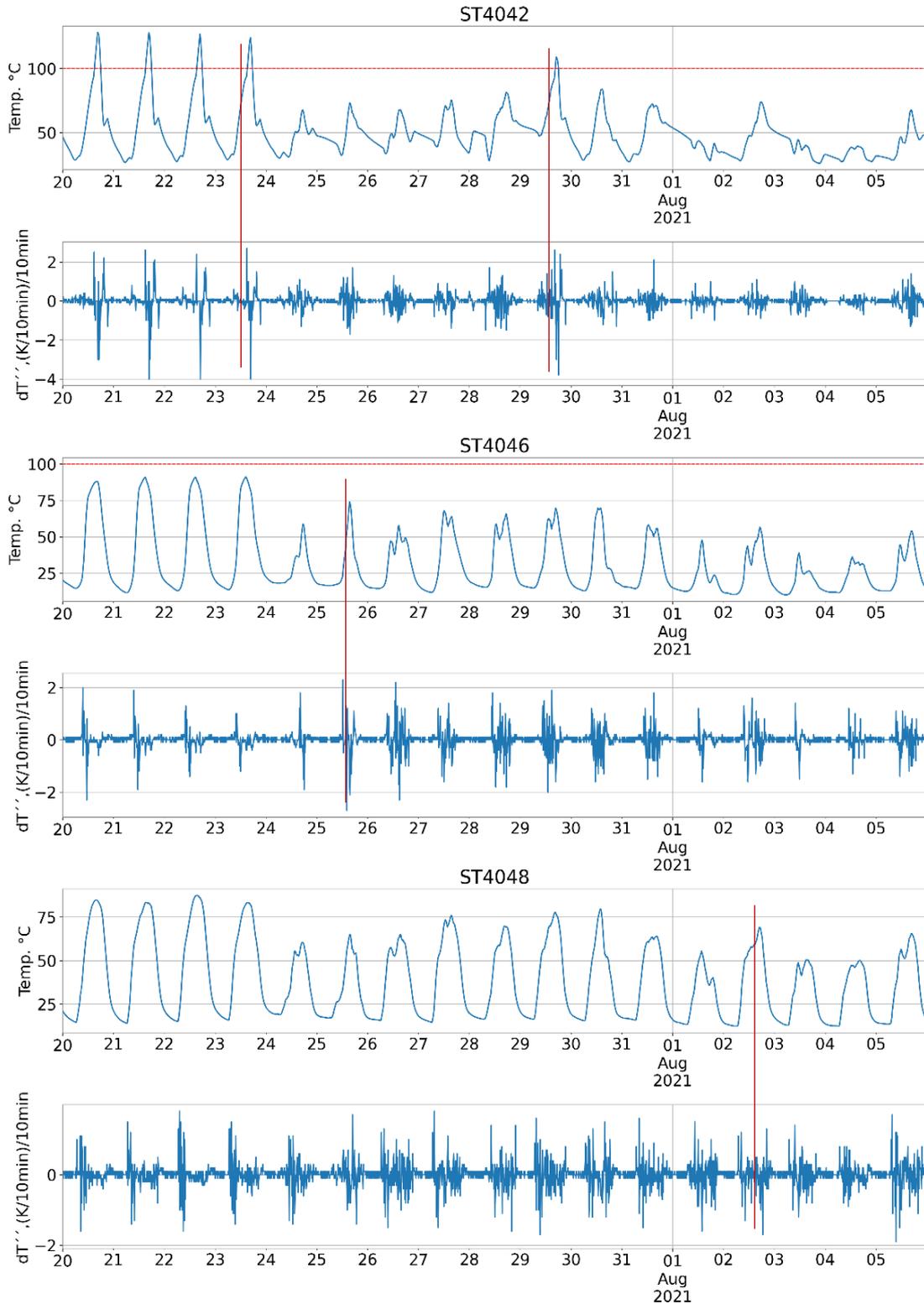


Abbildung 17 Sonnenkollektortemperatur von drei verschiedenen Sensoren und ihre jeweiligen zweiten Ableitungen



### 3.4 Testing

Um den regelbasierten Algorithmus schnell und robust implementieren und testen zu können, wurde er in Python-Pandas unter Verwendung eines jupyter-Notebooks (<https://jupyter.org/>) entwickelt. Um die Implementierungszeit zu reduzieren, wurde ein sauberer Datensatz, der zu Beginn des Projekts mit erheblichem datentechnischem Aufwand in Node-RED vorbereitet wurde, direkt für die Analyse verwendet. Eine zusätzliche Datenbereinigung in Python wurde nicht durchgeführt. Es wurde ein Zeitraum vom 19.10.2020 bis zum 19.10.2021 verwendet. Dies war einer der letzten vollständigen Datensätze, die zur Verfügung standen.

Die Tests wurden in **zwei Phasen** durchgeführt. In der ersten Phase wurden zwei Wochen aus jeder Jahreszeit ausgewählt und dann die Tage von Interesse innerhalb dieser Wochen bestimmt. Diese Tage umfassten alle typischen Vorfälle, die für die KPI-Erkennung von Interesse waren. Der RBA wurde an diesen Tagen durchgeführt und seine Wirksamkeit festgestellt. In der zweiten Phase wurde die RBA dann für den gesamten Zeitraum über Zeitspannen von jeweils drei Tagen durchgeführt, da in dieser Zeitspanne keine wesentlichen Änderungen der Wetter- oder Betriebsbedingungen erwartet wurden. *Abbildung 18* zeigt zwei Wochen in der Sommersaison, die für die drei Sensoren ausgewählt wurden. Innerhalb dieses Zeitrahmens sind Tage mit gutem Betrieb, Tage mit Überhitzung über 100 °C, Tage mit abrupten Temperaturschwankungen und Tage ohne Betrieb dargestellt.

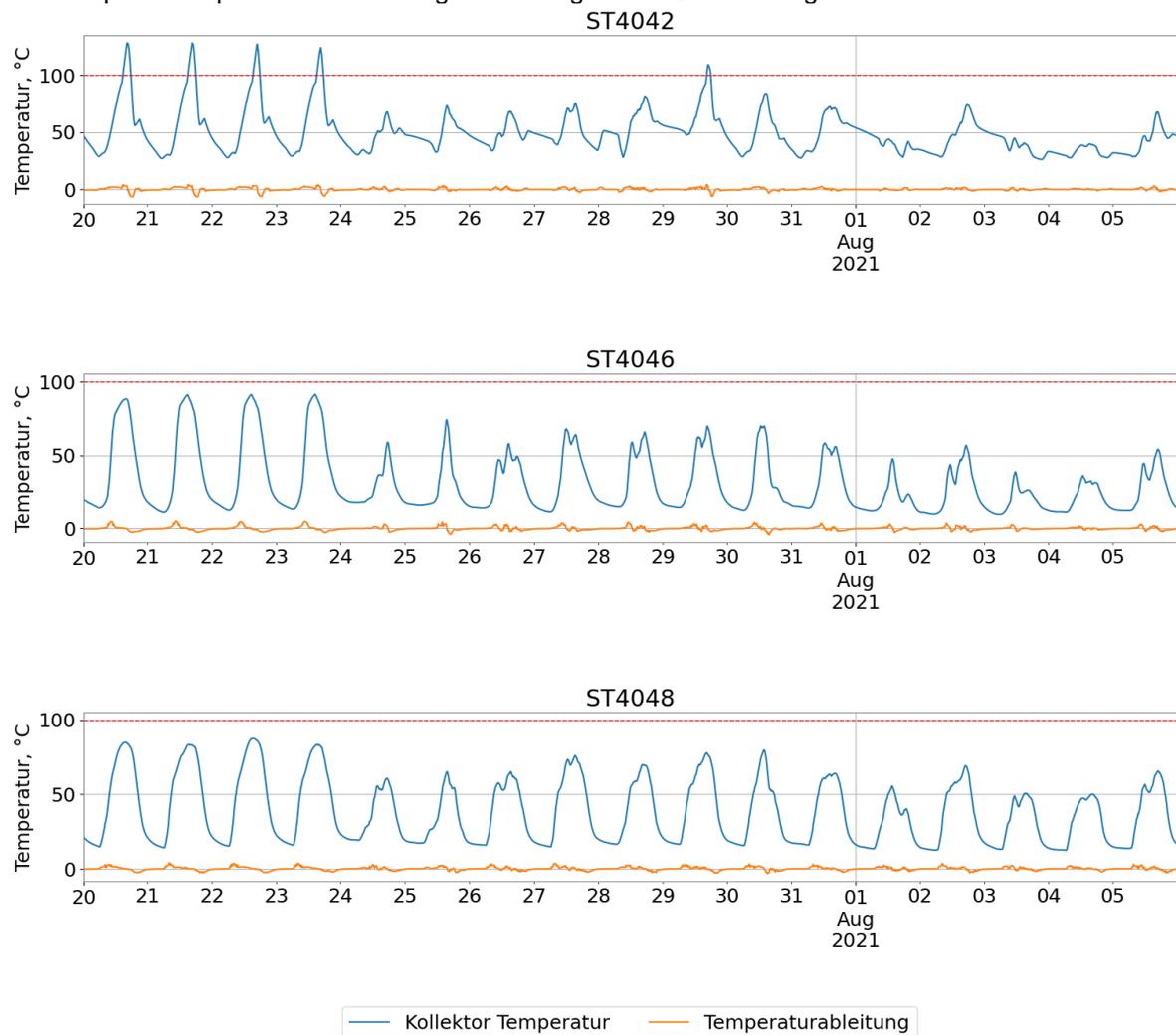


Abbildung 18 Beispiel eines zweiwöchigen Sommerdatensatzes, der zur Validierung des RBA verwendet wurde



### 3.4.1 Ergebnisse der KPI-Analyse für die KPIs N1...N4

Abbildung 19 unten zeigt die Ergebnisse des RBA für die KPIs N1...N4 am 22. Juli. Der anfängliche Anstieg der Vorlauftemperatur wird mit N1 korrekt identifiziert und der Pumpenstart nach einem vordefinierten Zeitparameter mit N2. Die Beendigung des Anlagenbetriebs wird in beiden Diagrammen mit N3 dargestellt, was die Möglichkeit bietet, die Gesamtbetriebsstunden zu schätzen, die durch den schattierten Bereich gekennzeichnet sind. N4 stellt die während des Tages erreichte Höchsttemperatur dar.

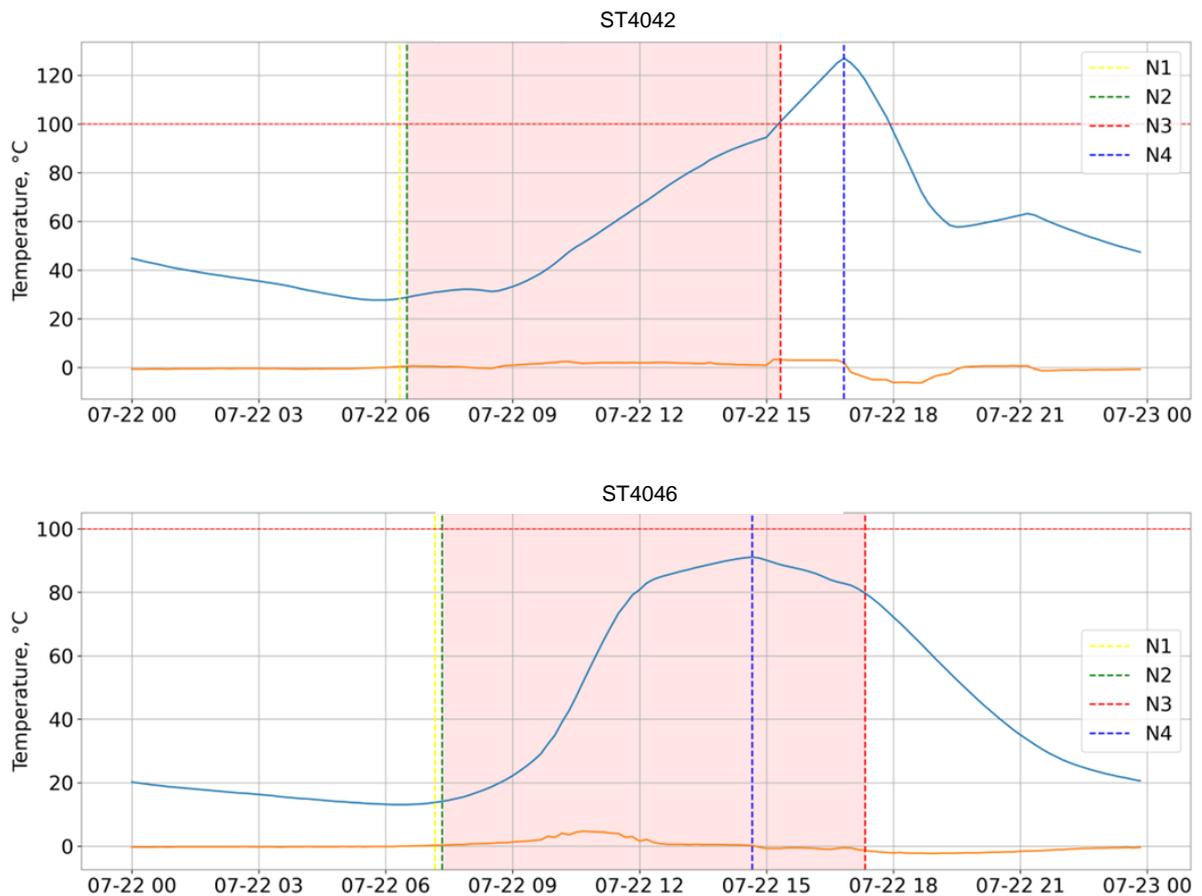


Abbildung 19 Beispiele für die Ergebnisse der RBA zur Identifizierung der KPIs N1..N4

Verschiedene andere Ereignisse mit komplexen physikalischen Phänomenen konnten mit dem RBA identifiziert werden. Ein solches Phänomen war das Auftreten mehrerer Spitzenwerte an einem einzigen Tag aufgrund von Bewölkung oder hydraulischen Faktoren. Die RBA war in der Lage, den Pumpenbetrieb auch an diesen Tagen mit akzeptabler Genauigkeit zu identifizieren, wobei die Abstimmung der Parameter eine wichtige Rolle spielte, um diese Genauigkeit zu erreichen.

Ein Beispiel für dieses Phänomen und die Auswirkung von Parametervariationen ist in der *Abbildung 20* unten dargestellt. Hier ist zu erkennen, dass die Temperatur des Sensors zwischen 12:00 und 15:00 Uhr vorübergehend abnimmt, bevor sie nach 15:00 Uhr wieder ansteigt. In der obigen Abbildung führt die Parametrierung dazu, dass N3 bereits nach dem ersten Temperaturabfall (erste Spitze) aktiviert wird. Es ist jedoch zu erkennen, dass die Anlage weiter in Betrieb war. Durch Optimierung des N3-Parameters war es möglich, die zweite Spitze zu erfassen, was eine bessere Einschätzung des Anlagenbetriebs ermöglicht.

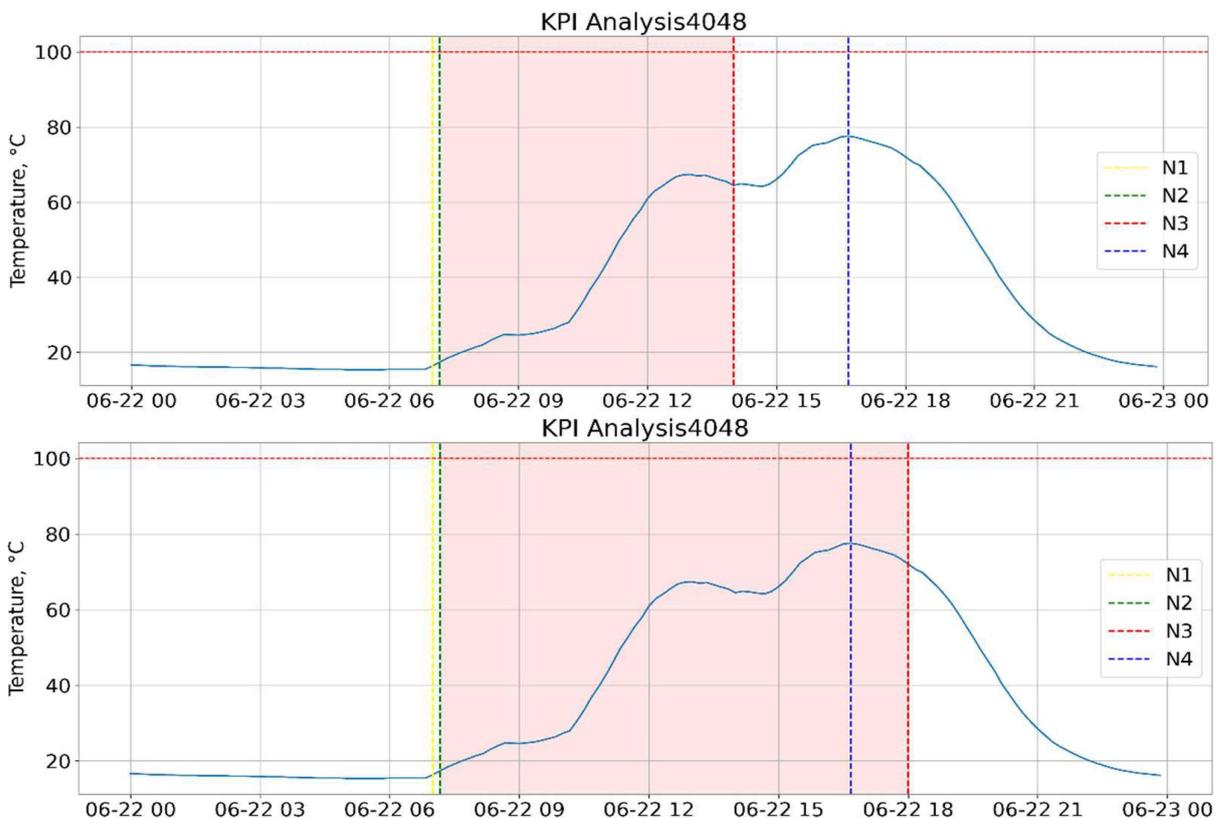


Abbildung 20 Beispiel für eine Parametervariation bei der KPI-Identifizierung

Dieser Ansatz hat seine Grenzen. Ein bestimmter Satz von Parametern kann die Genauigkeit für eine bestimmte Solaranlage erhöhen, aber er kann die Erhöhung der Genauigkeit nicht für alle Anlagen garantieren. Für künftige Arbeiten wurden hier zwei Möglichkeiten diskutiert:

1. Implementierung verschiedener Parametersätze für verschiedene Anlagen auf der Grundlage von Feldtests der RBA.
2. Einbeziehung von Informationen über die Bewölkung in die RBA-Analyse und nicht nur zur Filterung des Datensatzes.

### 3.4.2 Ergebnisse der KPI-Analyse für die KPIs N5...N7

Die folgende *Abbildung 21* zeigt die Ergebnisse des RBA für die Analyse der KPIs N5...N7. Als Beispiel sind die Daten verschiedener Sensoren für die Tage dargestellt, an denen die verschiedenen KPIs aktiviert wurden. Wie zu sehen ist, werden alle KPIs entsprechend ihrer Logik aktiviert und auf den Zeitpunkt gesetzt, an dem die Analyse nachts beginnt (19:00 Uhr). Im ersten Beispiel wurde keine Nachtauskühlung beobachtet, was durch die Aktivierung des KPI N5 gekennzeichnet ist. Hier kühlt die Temperatur natürlicherweise auf die Aussentemperatur (bzw. auf die Kellertemperatur) ab. Im zweiten Beispiel wurde jedoch für das Gerät *ST12108* eine Nachtauskühlung aufgrund eines defekten Rückschlagventils festgestellt, was zu erhöhten Messwerten führt, die jedoch langsam wieder abnehmen. Weiterhin wurde am Beispiel von *ST12091* am 30.07.2021 eine Nachtauskühlung durch unnötigen Pumpenbetrieb festgestellt. Hier bleibt die gemessene Vorlauftemperatur mit minimalen Schwankungen konstant warm (über 35 °C). Im letzten Bild wurde am Beispiel von *ST12123* am 08.06.2023 eine rohrinterne Zirkulation gezeigt. Der Grund für die Platzierung der inaktiven KPIs am Anfang des Analysezeitraums (links im Diagramm) liegt in der Struktur des Codes im Entwicklungssystem.

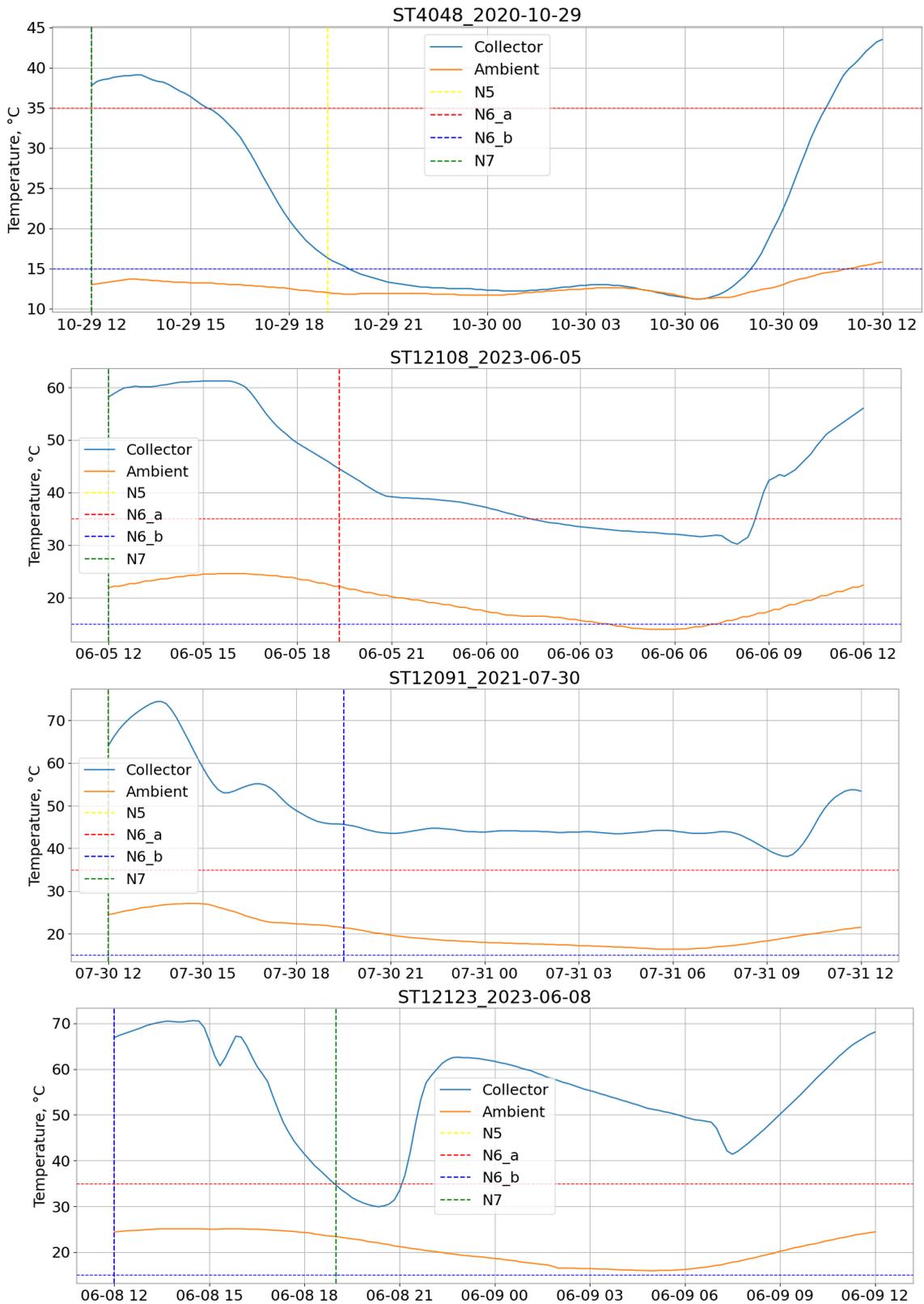


Abbildung 21 Beispiele für die Ergebnisse der RBA zur Identifizierung der KPIs N5..N7



Diese Ergebnisse beschreiben den aktuellen physikalischen Zustand einer Anlage anhand eines Tagesdatensatzes. Jedem KPI sind eines oder mehrere physikalische Parameter als Attribute zugeordnet. Diese ermöglichen die sinnvolle Interpretation der KPI. Diese Attribute und Labels und ihre Entwicklung werden im folgenden Abschnitt beschrieben.

### 3.5 Labels und Attributes

Zusätzlich zu den KPIs sollten auch wichtige Informationen über den Status der Solarthermieanlage, das Wetter, die Qualität der Daten und andere identifizierte Aspekte gespeichert werden. Diese Informationen sollen in Beziehung zu den KPIs stehen. Um dies zu erreichen, wurden verschiedene Attribute definiert, die denselben Zeitstempel wie die KPI erhalten. Auf diese Weise wurde ein gelabelter Datensatz wie unten beschrieben erstellt.

Zunächst wurde eine vollständige Liste der physischen Parameter erstellt, die als Attribute definiert werden sollen (siehe *Tabelle 2* unten).

Tabelle 2 Physikalische Attribute zu den KPIs.

Attribut	Beschreibung
<b>N</b>	Numerischer Wert 1 bis 8 zur Visualisierung der KPIs
<b>cloud</b>	Mittlerer Bewölkungsgrad des Tages
<b>girr</b>	Mittlere Bestrahlungsstärke in der Horizontalen
<b>minwtempc</b>	Mittlere Aussentemperatur
<b>pIndex</b>	Anzahl der 10min Intervalle ab Mitternacht bis zum Auftreten von N1
<b>pIndexN1</b>	Anzahl der 10min Intervalle ab N1 bis vor die abrupte Temperaturänderung
<b>pTdN1_N3</b>	Anzahl der 10min Intervalle zwischen N1 und N3
<b>pindexwith</b>	Anzahl der 10min Intervalle ab N2 bis zum Ende der monotonen Steigung
<b>setup</b>	Ort des Sensors am Kollektorfeld (1), bzw. im Technikraum (0).
<b>sumiss</b>	Anzahl nichtvorhandener Datenpunkte an dem zu untersuchenden Tag
<b>tempc</b>	Temperatur des Sensors bei Ansprechen des entsprechenden KPI
<b>dertempc</b>	Ableitung der Sensortemperatur bei Ansprechen des entsprechenden KPI
<b>tempcN4</b>	Temperatur unmittelbar vor einer raschen Änderung der Ableitung
<b>delta_t</b>	Temperatur – Differenz zwischen minimalem Sensortemperatur und minimaler Aussentemperatur bei Nacht.

Diese Attribute werden für die spätere Verwendung in der Datenbank gespeichert. Anschliessend wird der Datensatz gemäss *Abschnitt 3.1* aufbereitet, sodass für die Auswertung ein Tagesdatensatz mit 144 äquidistanten Werten für die Sensortemperatur, die Umgebungstemperatur und die Bestrahlungsstärke vorliegt.

Die Matrix definiert ein mathematisches Verfahren, das auf eine Tagesdatensatz der entsprechenden Zeitreihen angewendet wird. Die Anwendung des Verfahrens liefert den Wert «True», falls der entsprechende KPI zutrifft, ansonsten «False». Die mit KPI bezeichnete Spalte enthält die vier KPI1 bis KPI4, die bereits im vorangegangenen Projekt verwendet wurden, sowie sieben weitere KPIs, die mit N1 bis N7 bezeichnet sind. Diese KPIs sind als Labels von 1 bis 11 nummeriert. KPI und Label werden



häufig synonym verwendet. Streng genommen enthalten diese Begriffe unterschiedliche Aussagen. Der KPI beschreibt ein messbares, durch Zahlenwerte darstellbares Ereignis. Das zugehörige Label bezeichnet den Datensatz, bei dem der KPI auftritt. *Tabelle 3* zeigt das Resultat dieser Datenmodelle bzw. der Datenbank in kompakter Form. In dieser Tabelle ist auch zu sehen, welche Attribute den KPI zugeordnet sind.

Tabelle 3 Beispiel für die Einrichtung der Datenbank nach «Cleaning» und «Labelling» mit KPIs

Data_const			Data_day					Data_series					
Device	Grösse	Date	Grösse	Grösse	KPI + Attribute A0..A4		Daytime	n	Grösse	Grösse	KPI + Attribute N1..NX		
-	Einheit	yyyy-mm-dd	Einheit	Einheit	Einheiten		min	-	Einheit	Einheit	Einheiten		
123		2022-08-18					0	1					
							10	2					
							⋮	⋮					
							1380	144					
123		2022-08-19					0	10					
							10	2					
							⋮	⋮					
							1380	144					
123		2022-08-20					10	2					

### 3.6 Alarms

Treten in einem Tagesdatensatz gewisse KPIs einzeln oder andere KPIs gemeinsam auf, so wird dies als Fehler interpretiert. Um jedoch die Art und Intensität des Alarms zu bestimmen, sind weitere Informationen erforderlich. Dies geschieht anhand der gelabelten Datensätze und ihrer Attribute, wie im vorherigen Abschnitt beschrieben. *Tabelle 4* zeigt diese Beziehung in Form einer Matrix.

Die Spalte «Label» zeigt den Wert des Datensatzes an, der dem KPI entspricht. Entweder ein gelabelter Datensatz und die dazugehörigen Attribute oder eine Kombination aus mehreren Datensätzen und deren Attribute werden verwendet, um einen Alarm zu erzeugen und dessen Intensität zu bestimmen. Diese logische Kombination wird in der Matrix unter der Spalte «vertikale Auswertung» angezeigt. Die Alarms sind in vier Hauptkategorien unterteilt: «Datenausfall», «Batterie leer», «Stagnation», und «Nachtauskühlung». Diese sind in Unterkategorien (a..e) und fünf Intensitätsstufen unterteilt, wie in *Tabelle 5* beschrieben.

Tabelle 4 Zusammenhänge zwischen KPIs, Labels, Attribute und des daraus resultierenden Alarms für den Anwender

KPI	vereinfachte KPI-Beschreibung	Label	Vertikale Auswertung für Alarms ↓											
KPI1	Seit 24h keine neuen Daten	1	X											
KPI2	Batteriespannung < 3.3 V	2		X										
KPI3	Kollektortemperatur > 100 °C	3			X	X								
KPI4	3 Tage hintereinander KPI3	4				X								
N1	Erster Temperaturanstieg	5					X	X	X					
N2	Pumpenstart	6							X					
N3	Rasche Steigungsänderung	7						X	X					
N4	Maximaltemperatur am Tag	8												
N5	Normalbetrieb ohne Nachtauskühlung	9									X			
N6	Nachtauskühlung (Schwerkraftzirkulation oder Umwälzpumpe)	10											X	
N7	Nachtauskühlung (Rohrinterne)	11												X
Alarm Kategorie→					a	b	c	d	e			a	b	
			Dat	Bat	Stagnation							NK		



Tabelle 5 Alarmbeschreibung und Alarmstufen

Alarm Kategorie	Anwendung	Alarmstufe
<b>Datenausfall</b>	<u>Verbindung und Stabilität</u> von System prüfen	Hoch
<b>Batterie leer</b>	Meldung für <u>Batteriewechsel</u>	NA
<b>Stagnation-a</b>	<u>Erhitzung</u> : Anzahl Stunden von Sonnenaufgang bis Kollektortemperatur > 100 °C	Niedrig bis Hoch in fünf Stufen, je nachdem, wie schnell der Kollektor überhitzt
<b>Stagnation-b</b>	<u>Erhitzung</u> : Wenn 3 Tage hintereinander mind. einmal 100 °C erreicht wurde dann die Temperatur in der Solaranlage erreicht regelmässig über 100 °C.	Hoch
<b>Stagnation-c</b>	In Kombination mit Einstrahlungsdaten. Wenn trotz Einstrahlung kein N1 dann <u>Funktionsstörung</u>	Hoch
<b>Stagnation-d</b>	In Kombination mit Einstrahlungsdaten. Andere Attribute: Zeitstempel für N1 und N3 und Temp. bei N3. Wenn trotz Einstrahlung nur wenig Zeit zwischen N1 und N3 dann <u>Funktionsstörung</u>	Niedrig bis Hoch in fünf Stufen, basierend auf der Laufzeit zwischen N1 und N3
<b>Stagnation-e</b>	In Kombination mit Einstrahlungsdaten. Andere Attribute: Zeitstempel für N1, N2, und N3 und Temp. bei N3. Wenn trotz Einstrahlung nur wenig Zeit zwischen N1 und N3 obwohl N2 würde detektiert dann <u>Optimierungspotenzial</u>	Niedrig bis Hoch in fünf Stufen, basierend auf der Laufzeit zwischen N1 und N3
<b>Nachtauskühlung-a</b>	<u>Optimierungspotenzial</u> : Regelung oder Rückschlagventil	Niedrig bis Hoch in fünf Stufen, basierend auf der Höhe der Temperaturunterschiede zwischen Kollektortemperatur und Umgebungstemperatur
<b>Nachtauskühlung-b</b>	<u>Optimierungspotenzial</u> : Speicheraus Kühlung	Hoch

Die Erzeugung eines Alarms der Kategorie «Stagnation» wird anhand des folgenden Beispiels erläutert:

Wenn die KPIs N1 und N3 am Tag der Auswertung aktiviert sind, dann werden die entsprechenden Datensätze mit Label 5 und Label 7 ausgewertet, um zu entscheiden, ob ein Alarm Stagnation-d erzeugt werden soll oder nicht. Dabei werden die Zeitstempel für N1 und N3, die Temperatur an N3 und der Tagesdurchschnitt der globalen Einstrahlung als Attribute verwendet. Wenn der rasche Temperaturanstieg kurz nach dem Pumpenstart auftritt, ist die Alarmstufe höher, und umgekehrt ist die Alarmstufe niedriger.



## 3.7 Implementierung

Für die Implementierung des Algorithmus zur Definition der KPIs und zur Generierung von Alarmen wurden im Rahmen des Projekts zwei Möglichkeiten entwickelt.

### 3.7.1 Implementierung in «Node-Red»

Im Vorgängerprojekt wurde der Algorithmus im Node-Red-Framework umgesetzt. Dementsprechend wurde im aktuellen Projekt das Framework erweitert, um die verbesserte RBA zu integrieren. Hierzu wurden die notwendigen «Node-Red Flows» hinzugefügt und die Parameter entsprechend den Ergebnissen der Python-Implementierung angepasst. Die Logik wurde mit Funktionen zur Charakterisierung des Temperatursignals unter Verwendung der Derivatanalyse implementiert. Die beiden Hauptfunktionen werden im Folgenden beschrieben:

1. Die Funktion **FindGreater** generiert aus dem Zeitreihensignal und einem vorgebbaren Schwellenwert ein Rechteckpuls, gespeichert als «Pulse». Dieser hat die Amplitude 1 bzw. -1, je nachdem, ob ein Schwellenwert über oder unterschritten wird. Die Länge des Pulses entspricht der Zeitdauer, während der eine Schwelle über- oder unterschritten ist. Diese sind analog zu den im Python-Code implementierten Parametern «Zeitperiode» und «Schwellenwert».
2. Die Funktion **CharacterisePulse** ermittelt die Dauer des Pulses.

Abbildung 22 zeigt das Beispiel eines Rechtecksignals, das durch Anwenden der Funktion *FindGreater* und *CharacterisePulse* auf die Ableitung des Temperatursignals generiert wurde.

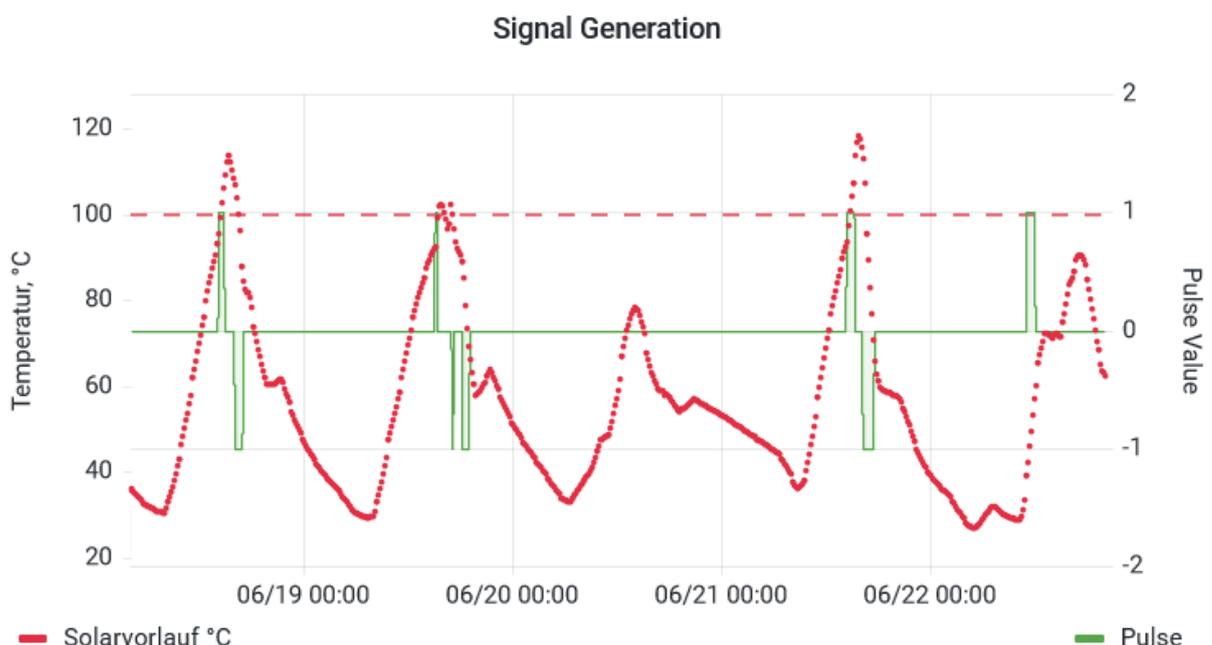


Abbildung 22 Beispiel eines Rechtecksignals «Pulse», generiert durch Anwenden der Funktion *FindGreater()* auf die Ableitung des Temperatursignals

Dieser Prozess wurde iterativ durchgeführt, was eine schrittweise Verbesserung und kontinuierliche Anpassung ermöglichte. Nach jedem Entwicklungsschritt wurde der zu diesem Zeitpunkt verfügbare Algorithmus getestet und auf das Produktivsystem übertragen. In der *Tabelle 6* sind die einzelnen Releases aufgelistet und die Neuerungen beschrieben:



Tabelle 6 Releases des regelbasierten Algorithmus im Node-Red

Nr.	Datum	Beschreibung
V0.0	16.06.22	Erster Release auf das Produktivsystem, mit N1, N2, N3 und N4; sowie den KPI1 bis 4 aus dem vorgängigen Projekt
V0.1	08.08.22	Zweiter Release, hinzufügen der KPIs N5, N6 und N7
V0.2	05.09.22	Ergänzen der fehlenden mathematischen Bibliotheken auf dem Produktivsystem. Überarbeitung des Datenmodells Nx mit Applikations-Experten, so dass die Beurteilung in Grafana effektiver vorgenommen werden kann.
V0.3	03.10.22	Überarbeitung der Softwarearchitektur auf dem Produktivsystem, so dass die FHNW selbständig eine Neuentwicklung auf dem Produktivsystem installieren kann. Hinzufügen eines Datenmodells Nxx für die KPIs N5, N6, und N7
V0.4	30.03.23	Verbesserung der bestehenden KPI-Logik für N1...N4 auf der Grundlage von visuellen Tests in Grafana
V0.5	03.04.23	Verbesserung der KPI-Logik auf der Grundlage der Python-Implementierung und der Parameterabstimmung. Korrektur der Daten-Filterung-Parameter für die Alarmgenerierung.
V1.0	13.06.23	Stabile Version des Node-Red Frameworks mit den KPIs N1 bis N4 und Alarmen für «Datenausfall», «Batterie leer» und «Stagnation». Installation im Produktivsystem und gesichert in einem Gitlab-Projekt.

Im Rahmen von regelmässigen Workshops wurde geprüft, welche weiteren Information über den Betriebszustand von Anlage aus den Signalverläufen und Umweltparametern ableiten lässt. Hierzu wurden sogenannte Dashboards in «Grafana» eingerichtet, die eine bequeme und rasche Visualisierung der Daten ermöglichen. Als Resultat wurde eine Metrik zur Analyse der Datensätze entwickelt. Die Anlage, die Zeitspanne, und die anzuzeigenden Grössen sind frei wählbar. Beispielsweise könnte auch die Umgebungstemperatur und der Bewölkungsgrad dargestellt werden. In der neuesten Version (V1.0) sind nur die KPIs N1-N4 enthalten (siehe *Tabelle 6*).

*Abbildung 23* zeigt den Ausschnitt aus dem Dashboard mit der Sensortemperatur (Solarvorlauf). Ebenfalls gezeigt sind die numerischen Indikatoren der KPIs N1..N4, denen der Wert «True» zugeordnet wurde. Für die Anzeige wird dieser boolesche Wert in die Ziffern 1..4 für den jeweiligen KPI umgewandelt und auf der sekundären Y-Achse angezeigt. Zusätzlich zu den Überwachungsdaten werden auch Informationen zu den Alarmen in tabellarischer Form angezeigt.

Dieses Dashboard wird durch Applikations-Experten verwendet, um den regelbasierten Code zu prüfen und um problematische Anlagen genauer zu untersuchen. Theoretisch ist bei diesem Schritt eine messbare Rückmeldung über die Genauigkeit des Algorithmus erforderlich, um einen MLA zu implementieren.



Abbildung 23 Beispiel für das von der FHNW implementierte Grafana Dashboard zur Anzeige verschiedener Monitoringdaten und Ergebnisse der RBA

### 3.7.2 Implementierung des RBA in «Flux»

Sowohl bei der Codeentwicklung im Projekt als auch beim Betrieb der Software auf dem Produktivsystem wurde festgestellt, dass sich Node-Red nur bedingt für komplexe Zeitreihenanalysen eignet. EZS hat daher entschieden, den RBA direkt in der zur Influx-Datenbank gehörenden Programmiersprache «Flux» zu programmieren. Im weiteren Projektverlauf wurde der RBA vollständig in das Produktivsystem implementiert. Es dient aktuell zur Überwachung von über 450 thermische Solaranlagen. Darunter sind allein 150 Solaranlagen, die im Contracting betrieben werden und 60 grosse Solaranlagen in Liegenschaften einer grösseren Pensionskasse.

Die FHNW unterstützte bei der Ausarbeitung und Implementierung der Algorithmen in «Flux».

Ein Beispiel für die Implementierung wird in der *Abbildung 24* gezeigt. Der Batteriestatus wird zusammen mit anderen Monitoring-Daten in verschiedenen Panels angezeigt. Zusätzlich sind auch die Alarme und ihre Stufen implementiert.



Abbildung 24 Beispiel für das von EZS implementierte Dashboard zur Anzeige verschiedener Überwachungsdaten

Die Informationen über die Betriebsstunden der Solaranlagen mit Hilfe der KPIs N1-N4 wurden in Form eines Energiereports implementiert, um die Energieproduktion und andere wichtige Parameter zu bewerten.

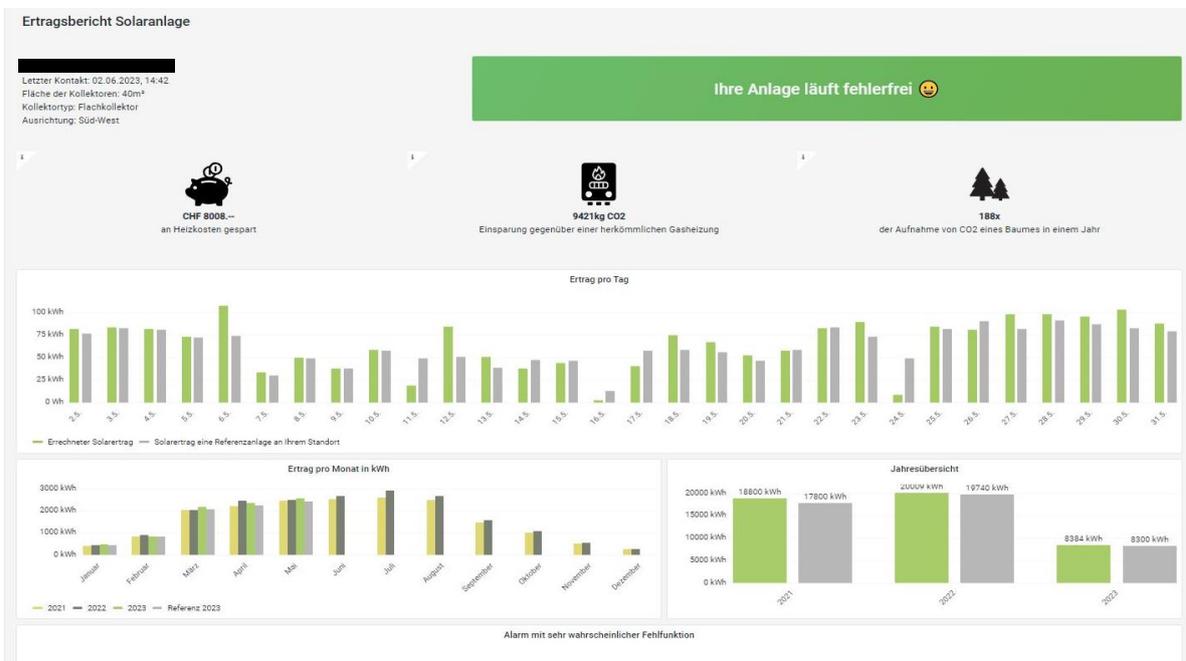


Abbildung 25 Beispiel für einen Bericht, der durch EZS für seine Endbenutzer unter Verwendung der RBA-Logik erstellt wurde



### 3.8 Bewertung der Ergebnisse

Eine Entwicklungsdatenbank wurde eingerichtet, auf welche die Messdaten in regelmässigen Abständen und in anonymisierter Form gespiegelt werden. Ein Algorithmus für die Aufbereitung der Daten für die Analyse wurde entwickelt, getestet und in Betrieb genommen. Dies erzeugt die so genannten KPIs. Aus diesen können direkt oder in Kombination Betriebszustände, Ereignisse, Fehlfunktionen und Kontextinformationen abgeleitet werden, beispielsweise der Zeitpunkt des Pumpenstarts, Stagnation, und Wärmeverlust durch Nachtauskühlung. Die Ergebnisse wurde durch manuelle Analyse auf ihre Genauigkeit überprüft. Die Analysen sind in der folgenden *Tabelle 7* zusammengefasst. Sie basieren sowohl auf der quantitativen Analyse der Ergebnisse für die in der Studie verwendeten Datensätze als auch auf dem qualitativen Feedback von EZS.

Tabelle 7 Ergebnisse zur Genauigkeit des RBA

Fehler	Zustand (für Ertragsabschätzung)	Ziel - Quantitativ Sicherheit %	Quantitativ Sicherheit %	Qualitativ Feedback
	<b>Pumpe ein/aus</b>	90	98	Sehr gut
<b>Nachtauskühlung des Speichers</b>		95	86	Zufrieden
<b>Stagnation durch Fehlfunktion</b>		90	93	Sehr gut
<b>Konkurrierende Zusatzheizung</b>		90	--	Zufrieden

Die Treffsicherheit des RBA ist erstaunlich gut. Er liefert auch bei der Nachtauskühlung, für deren Analyse ursprünglich ein MLA vorgesehen war, hinreichend zuverlässige Aussagen. Die etwas geringere Genauigkeit bei der Identifizierung der Nachtauskühlung ist auf die Tatsache zurückzuführen, dass es drei mögliche Gründe für eine solche Nachtauskühlung gibt: Nachtauskühlung durch Pumpenbetrieb, durch Schwerkraftzirkulation infolge undichten Rückschlagventils und durch rohrinterne Zirkulation infolge einer ungenügenden oder fehlenden Siphonierung. Zudem liefert die Datenbank keine absolut zuverlässige Auskunft über den Einbauort des Sensors. Daraus ergibt sich eine weitere Unsicherheit. Einerseits war die Glättung der Daten durch den Moving-Average Filter erforderlich, um den RBA überhaupt anwenden zu können. Andererseits führte diese Glättung zum Verlust wertvoller Merkmale, die Hinweise auf Ursachen geben könnten.

Die Orientierung des Kollektors war in den Datensätzen nicht eindeutig verfügbar, und es wurde vorgeschlagen, sie anhand der Kollektortemperatur zu schätzen. Dies würde auf der Grundlage der Zeitspanne zwischen dem ausgewerteten Pumpenstart und der Tageszeit erfolgen. Ein früher Start deutet auf eine östliche Orientierung und ein später Start auf eine westliche Orientierung hin. Eine wesentliche Einschränkung dieses Ansatzes war die fehlende Information über die Speichertemperatur, die ebenfalls den Pumpenstart beeinflusst. Ausserdem wäre diese Schätzung der Orientierung auf der Grundlage des Pumpenstarts, der wiederum auf der Grundlage der Kollektortemperatur berechnet wird, eine zusätzliche Analyse, die keinen vertretbaren Nutzen für die erhöhte Komplexität in der Logik bringen würde. Dies zeigt sich auch in den zufriedenstellenden Ergebnissen bei der Identifizierung von Fehlfunktionen, die direkt auf der Kollektortemperatur basieren. Es kann jedoch angenommen werden, dass die Information über die Orientierung die Schätzung des solarthermischen Ertrags verbessert und daher einbezogen werden sollte, wenn Informationen über die Orientierung im Voraus verfügbar sind. Nichtsdestotrotz wurden im Rahmen dieses Projekts zufriedenstellende Ergebnisse bei der Ertragsschätzung allein auf der Grundlage des Pumpenbetriebs erzielt.



Die Entscheidung, die Technologie von «Node-Red» auf «Flux» zu wechseln und das Know-how der Projektpartner zu erweitern, führte insgesamt zu einer verbesserten Anpassungsfähigkeit und Effizienz des RBA. Die ausführliche Dokumentation und die Unterstützung durch die FHNW ermöglichte diesen nahtlosen Übergang und trug zur erfolgreichen Umsetzung des Projekts bei.

Die wichtigsten Erkenntnisse sind: Zuverlässige Daten liefern wichtige Einblicke, ermöglichen frühzeitige Problemerkennung und fördern die kontinuierliche Verbesserung von Anlagen im Betrieb. Ohne solide Daten drohen fehlerhafte Analysen und suboptimale Entscheidungen. Eine gute Wahl der Datenquelle ist daher unerlässlich.

## 4 Einsatz maschinellen Lernens

Die erfolgreiche Implementierung von Tools für maschinelles Lernen im Rahmen dieses Projekts stösst auf mehrere Herausforderungen.

Die Auswahl der Daten für die Entwicklung des RBA erfolgten manuell und auf der Basis von Expertenwissen. Aufgrund des manuellen Prozessschrittes bei der Auswahl der Daten war es nicht möglich, eine ausreichende Anzahl Datensätze für das Training und die Validierung eines MLA zu labeln, was die Anwendung von Techniken des «Supervised Learning» zur Identifizierung der zuvor beschriebenen KPIs erschwerte. Darüber hinaus ist das Fehlen von menschlichem Feedback für jeden Tag eines jeden Sensors unpraktisch, insbesondere bei der für die Zukunft erwarteten, ständig steigenden Anzahl von Sensorinstallationen. Dieses fehlende Feedback verhindert die Bewertung eines jeden Modells und seine eventuelle Verbesserung. Das ursprüngliche Konzept, sich für das Training des MLA ausschliesslich auf die Ergebnisse der RBA zu verlassen, erwies sich ebenfalls als grundlegend fehlerhaft, da die theoretische Genauigkeit des MLA auf die Genauigkeit der RBA beschränkt wäre.

Diese Herausforderungen werden auch in der neueren wissenschaftlichen Literatur über Störungsarten in solarthermischen Anlagen und die Implementierung von datengesteuerten Techniken zu deren Erkennung diskutiert [13-14]. Die von Feier et al. [14] demonstrierte Implementierung unter Verwendung von Random Forest MLA erfordert jedoch einen grossen Datensatz (in ihrer Studie wurde 1 Jahr pro Anlage verwendet) und mehrere Sensoren für jede Anlage. Dies ist in unserem Fall unwahrscheinlich, da nur wenige Daten zur Verfügung stehen, vor allem, weil nur ein Sensor pro Anlage eingesetzt wird.

In Anbetracht des Mangels an Ressourcen für das persönliche Feedback muss der Implementierungsrahmen eine automatisierte Analyse ermöglichen, die alle Sensoren umfasst. Da ausserdem bekannt ist, dass sich jede Anlage von Natur aus anders verhält, wurde beschlossen, einen MLA so zu entwickeln, dass dieser in einer Schleife für jeden einzelnen Sensor ausgeführt werden kann, anstatt ein einziges Modell auf alle Sensoren anzuwenden. Dadurch wird der Einfluss des unterschiedlichen Anlagenverhaltens eliminiert und die dadurch bedingte Unschärfe vermieden.

Gestützt durch die Empfehlungen aus der Literatur und die oben genannten Herausforderungen wurde der Umfang der Implementierung von MLA zur Fehlererkennung wie folgt definiert:

- Zeitreihenprognose MLA aufgrund begrenzter Daten zur Klassifizierung von Fehler- und Nicht-Fehlertagen unter Verwendung eines anlagenspezifischen Modells
- Unüberwachtes Lernen zur Klassifizierung von Anlagentypen mit einem einzigen Algorithmus, der auf allen verfügbaren Daten trainiert wurde

### 4.1 Zeitreihenanalyse

Vor der Implementierung von Tools für maschinelles Lernen zur Vorhersage ist ein klares Verständnis des Datensatzes erforderlich. Mit Hilfe des Grafana-Dashboards wurde eine umfassende visuelle Analyse für alle Sensoren durchgeführt, um die typischen Muster zu verstehen, die für die



Fehlererkennung prognostiziert werden können, oder um die Attribute zu definieren, die für die Klassifizierung der Sensoren verwendet werden können. Es wurde eine Korrelationsmatrix zwischen den verschiedenen Attributen erstellt. Die folgende *Tabelle 8* zeigt zum Beispiel die Ergebnisse für ST4048.

Tabelle 8 Korrelationsmatrix für die verschiedenen Attribute im Datensatz, der für die Implementierung eines MLA verwendet wird

	<b>CollectorVorlauf</b>	<b>Cloud_Cover</b>	<b>T_Ambient</b>	<b>Global_Irradiation</b>
<b>CollectorVorlauf</b>	1.000000	-0.291566	0.758347	0.865909
<b>Cloud_Cover</b>	-0.291566	1.000000	-0.154176	-0.345686
<b>T_Ambient</b>	0.758347	-0.154176	1.000000	0.510468
<b>Global_Irradiation</b>	0.865909	-0.345686	0.510468	1.000000

Es zeigt sich, dass eine starke positive lineare Korrelation mit einem Koeffizienten von 0,86 zwischen der Vorlaufemperatur und der Sonneneinstrahlung besteht. Es ist offensichtlich, dass die Zunahme/Abnahme der Sonneneinstrahlung eine direkte Ursache für die Zunahme/Abnahme der Temperatur ist und daher in die Vorhersagemodelle einbezogen werden sollte.

Aus dem gesamten Datensatz wurden drei aufeinanderfolgende störungsfreie Tage identifiziert und eine Zeitreihenanalyse durchgeführt, indem die Trainingsdatenreihen zerlegt wurden. Das «statsmodels»<sup>4</sup> Library für Python wurde implementiert. Wie unten zu sehen ist, kann die ursprüngliche Zeitreihe additiv in ihre Hauptkomponenten Trend, Saisonalität und Residuen zerlegt werden. Das Temperaturprofil für die angegebenen Tage ist in der ersten Abbildung dargestellt. Während der Betriebsstunden steigen die Temperaturen von etwa 25 °C auf 90 °C und fallen in den Abendstunden wieder ab. Die Trendkomponente fängt den konstanten Teil dieses Verhaltens ein und empfiehlt ~ 50 °C als Grössenordnung. Das restliche Muster wird versucht, mit der Saisonalitätskomponente zu erfassen, die das «tägliche Glockenkurvenverhalten» einer solarthermischen Anlage darstellt. Die verbleibenden Werte nach Hinzufügen der Trend- und Saisonalitätskomponente, die vom Modell nicht erfasst werden konnten, werden in den Residuen angezeigt.

<sup>4</sup> <https://www.statsmodels.org/dev/about.html>

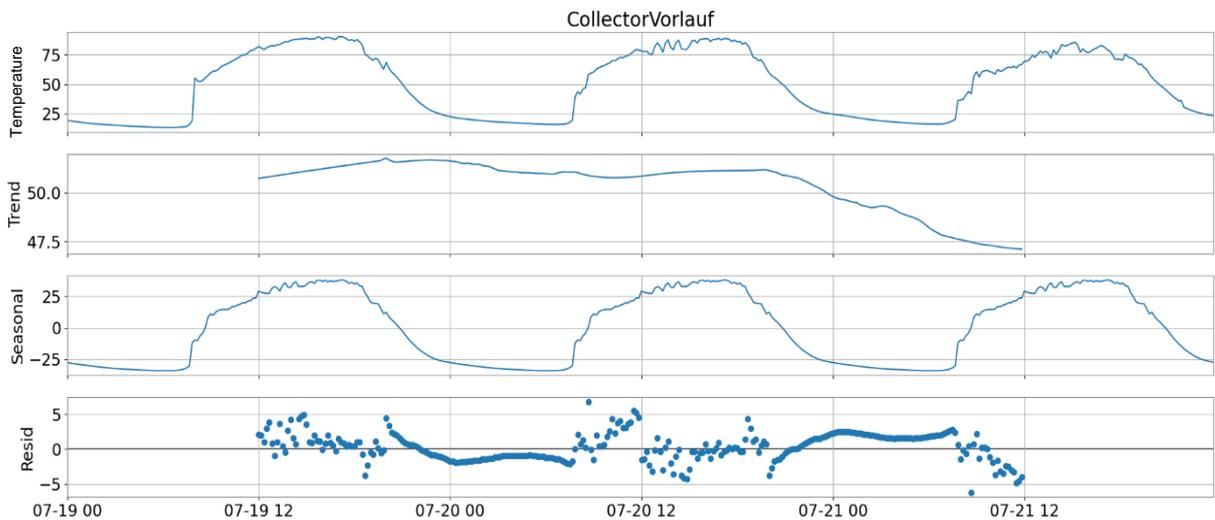


Abbildung 26 Beispiel für die Zerlegung einer Zeitreihe in ihre drei Hauptkomponenten: Trend, Saisonalität und Residuen

## 4.2 Exponentielle Glättung

Die exponentielle Glättung ist eine weit verbreitete Technik der Zeitreihenanalyse, bei der vergangenen Beobachtungen eine exponentiell abnehmende Gewichtung zugewiesen wird. Diese Methode priorisiert die jüngsten Daten und bietet einen reaktionsfähigen und anpassungsfähigen Ansatz zur Erfassung von Trends und Mustern in Zeitreihendaten. Die Prognose wird als gewichtete Summe vergangener Beobachtungen berechnet, wobei der Gewichtungsfaktor exponentiell abnimmt, je weiter die Datenpunkte in die Vergangenheit zurückreichen. Das «Holt-Winters Seasonal Exponential Smoothing» wurde aus der statsmodels Bibliothek implementiert [15].

Wie in der *Abbildung 27* unten gezeigt, werden drei Tage aus der Vergangenheit verwendet, um ein exponentielles Glättungsmodell für die Vorhersage des erwarteten Musters in den nächsten drei Tagen zu erstellen. Dies ist eine Übertreibung bzw. eine gewaltige Extrapolation der Fähigkeiten des Modells, dient aber nur zu Demonstrationszwecken. Im Produkktivsystem sollte nur ein Tag in die Zukunft prognostiziert werden. Es zeigt sich hier, dass die beiden Tage der Vorhersage am 23. und 24. Juli die notwendigen Voraussetzungen erfüllen, um als störungsfreie Tage identifiziert zu werden. Die Vorhersage für den dritten Tag, d.h. den 25. Juli, weicht jedoch stärker von den gemessenen Daten ab. Dies wird quantifiziert durch die Überschreitung eines Schwellenwertes für den «Root Mean Squared Error» (RMSE) zwischen gemessenen und vorhergesagten Daten. Der RMSE wurde gewählt, da er eine höhere Strafe für signifikante Abweichungen vorsieht. Daher wird er als Fehlertag gekennzeichnet, der von einem Menschen im Detail untersucht werden muss. Die Abweichungen von der Vorhersage sind in diesem Fall höchstwahrscheinlich auf die Bewölkung und die unregelmässige Sonneneinstrahlung zurückzuführen, wie Sie an den Spitzen in den Messdaten erkennen können. Die Einschränkung der exponentiellen Glättungsmethode besteht darin, dass sie die Auswirkungen der Bewölkung oder der Sonneneinstrahlung auf die Kollektortemperatur nicht berücksichtigen kann. Daher wurde eine andere Methode für die Zeitreihenprognose gewählt, die es ermöglicht, vergangene Daten mit saisonalen Mustern und externen Parametern wie der Einstrahlung einzubeziehen. Dies ist mit der Methode «Seasonal Autoregressive Moving Average Exogenous» (SARIMAX) möglich und wird in nächsten Abschnitt beschrieben.

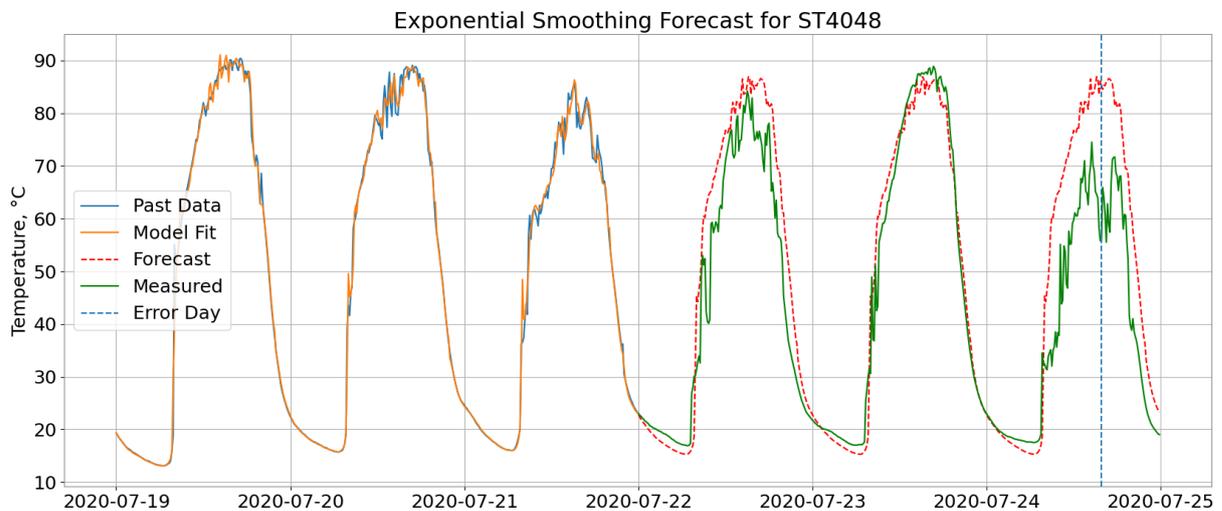


Abbildung 27 Beispiel einer prognosebasierten Fehlererkennung unter Verwendung der exponentiell glättenden Zeitreihenvorhersagemethode

### 4.3 SARIMAX

SARIMAX ist ein leistungsfähiges und flexibles Verfahren zur Analyse von Zeitreihen. Es wurde entwickelt, um Zeitreihendaten zu verarbeiten, die sowohl einen Trend als auch eine Saisonalität aufweisen, und gleichzeitig die Auswirkungen exogener oder externer Variablen zu berücksichtigen. Die autoregressive Komponente stellt die Beziehung zwischen einer Beobachtung und mehreren verzögerten Beobachtungen (frühere Zeitpunkte) dar, während die Komponente des gleitenden Durchschnitts die Beziehung zwischen einer Beobachtung und einem Restfehler aus einem Modell des gleitenden Durchschnitts, das auf verzögerte Beobachtungen angewendet wird, erfasst. Die Integrationskomponente ist notwendig, da die ursprüngliche Zeitreihe differenziert wird, um sie stationär zu machen, da solche Prognosemodelle nur auf eine stationäre Datenreihe angewendet werden können. Die tägliche Saisonalität in der solarthermischen Zeitreihe, wie in der obigen *Abbildung 26* dargestellt, kann mit der saisonalen Komponente erfasst werden. Die exogene Variable wäre in diesem Fall die globale Einstrahlung, die zur Verbesserung der Vorhersagegenauigkeit verwendet wurde, indem zusätzliche Informationen erfasst wurden, die die Zeitreihe beeinflussen. Die Modellparameter und die Bedeutung der exogenen Variable wurden durch eine Zeitreihenanalyse bestimmt, z.B. durch die Analyse von Autokorrelations- und partiellen Autokorrelationsfunktionen, und erforderten ein gewisses Mass an Versuch und Irrtum.

Wie im *Abbildung 28* zu sehen ist, ist das SARIMAX-Modell in der Lage, die Auswirkungen der globalen Einstrahlung zu erfassen und schätzt den letzten Tag des Vorhersagezeitraums besser als die exponentielle Glättung. Dieser wird nicht als fehlerfreier Tag markiert. Dadurch verringert sich die Anzahl der falsch-positiv Resultaten. Nur Tage mit Abweichungen, die nicht unbedingt auf die Bewölkung zurückzuführen sind, werden als Fehlertage identifiziert.

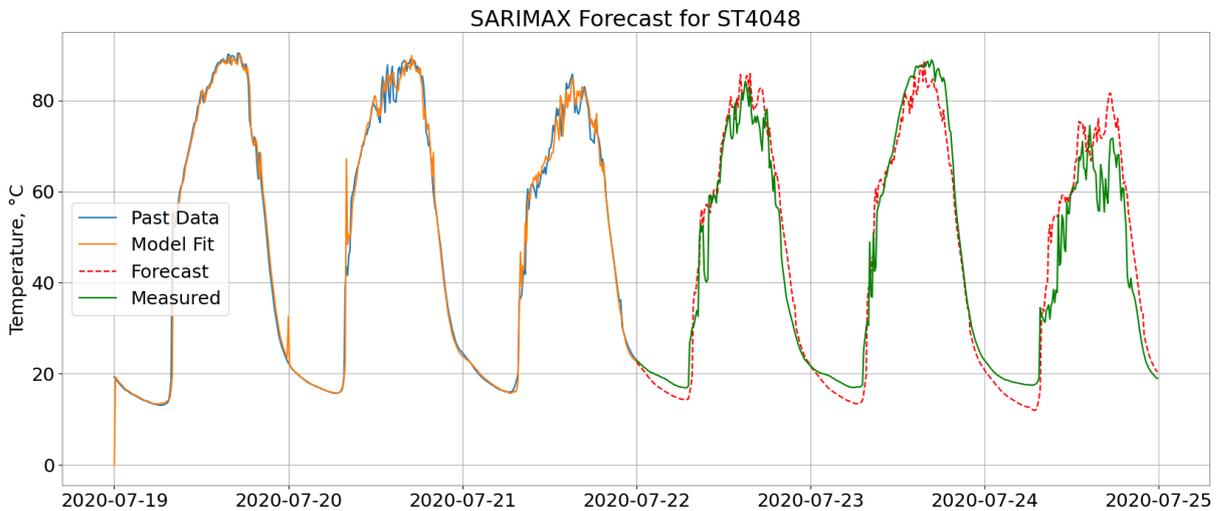


Abbildung 28 Beispiel für eine prognosegestützte Fehlererkennung mit der SARIMAX-Zeitreihenvorhersagemethode

#### 4.4 Bewertung der Ergebnisse

Die in diesem Abschnitt vorgestellten Ergebnisse beruhen auf einer vorläufigen MLA-Implementierung mit begrenzten Daten, Rechenressourcen und einem Mangel an kritischem Einblick in diese Methoden. Die Plausibilität dieser Methoden wurde jedoch bei der Diskussion der Ergebnisse im Projektteam festgestellt und es wird dringend empfohlen, diese Themen weiter zu verfolgen. Die wichtigsten Vorteile sind ein automatischer Algorithmus zur Fehlererkennung ohne menschliches Eingreifen zur Identifizierung oder Vorbereitung der Trainingsdaten bei jeder neuen Anlage oder bei der Nachrüstung einer bestehenden Anlage mit dem EZS-Überwachungssystem.

#### 4.5 Empfehlung für die Softwarestruktur zur Implementierung der Zeitreihenmethode

Obwohl jede der oben beschriebenen zeitreihenbasierten Methoden für die Fehlererkennung implementiert werden kann, wird die SARIMAX-Prognose aufgrund ihrer höheren Genauigkeit empfohlen. Wie in *Abbildung 29* dargestellt, könnte die Umsetzung wie folgt erfolgen:

Zu Beginn der Analyse definiert der Benutzer einen Trainingsdatensatz, der auf weniger als drei aufeinanderfolgenden störungsfreien Tagen basieren kann. Dieser Trainingsdatensatz wird verwendet, um ein SARIMAX-Modell anzupassen. Das SARIMAX-Modell wird verwendet, um eine Prognose für einen Tag in der Zukunft, den Zieltag, zu erstellen. Wenn die gemessenen Daten und die prognostizierten Daten für den Zieltag ähnlich sind, wird der Zieltag als störungsfreier Tag gelabelt, andernfalls als Störungstag. Der Trainingsdatensatz wird mit dem Ergebnis des Zieltages aktualisiert. In der nächsten Iteration, d.h. für den nächsten Zieltag, werden die aktualisierten Trainingsdaten verwendet. Es werden also immer die letzten drei störungsfreien Tage verwendet, um das SARIMAX-Modell anzupassen. Im realen Einsatz soll die Schleife nur einmal pro Tag und Sensor ausgeführt werden und der Benutzer muss nur einmal pro Sensor die Trainingsdaten definieren.

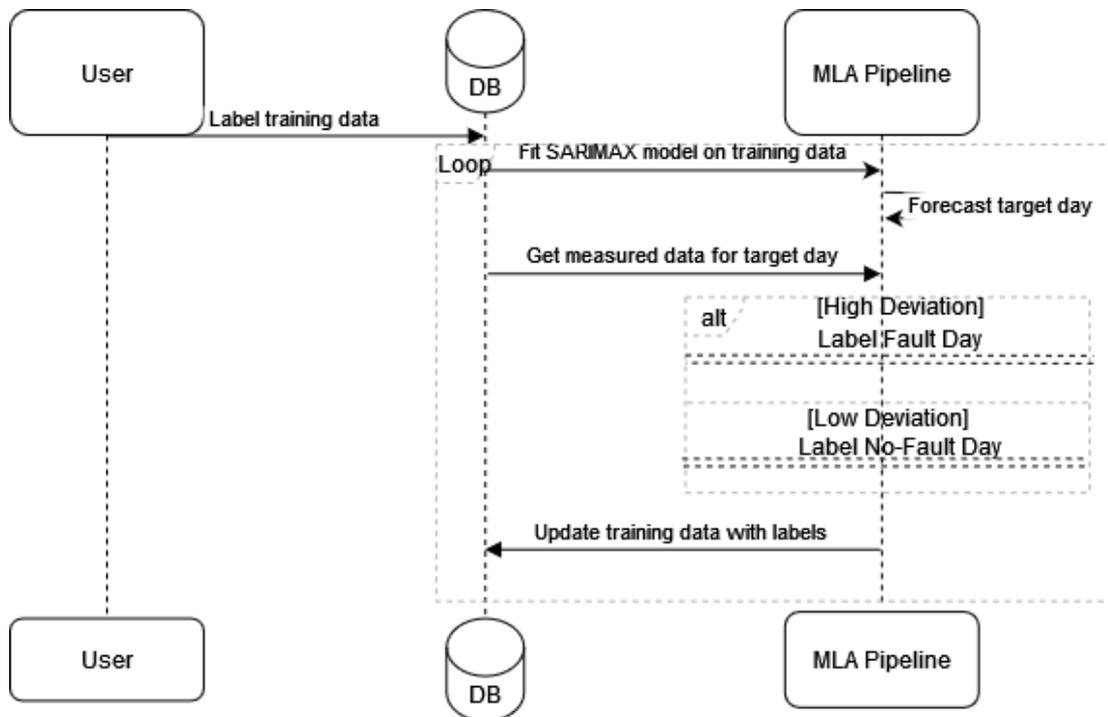


Abbildung 29 Beispielhafte Softwarestruktur zur Implementierung der MLA-Pipeline in das Produktivsystem von EZS

## 5 Ausblick

Solarthermische Anlagen liefern eine klimaneutrale Wärme, die vor Ort erzeugt wird und über Jahrzehnte hinweg wirtschaftlich stabil ist. Europäische Hersteller und regionale Experten können den gesamten Produktions-, Installations- und Betriebsprozess übernehmen. In Kombination mit einem geeigneten thermischen Speichersystem können diese Anlagen einen wesentlichen Teil des jährlichen Wärmebedarfs, insbesondere in den Sommermonaten, decken. Um Ihre Zuverlässigkeit und betriebliche Effizienz zu erhöhen, ist nicht nur eine verbesserte Planung und Dimensionierung der Systemkomponenten und der Hydraulik erforderlich, sondern auch eine kontinuierliche Überwachung und Verbesserung des Systembetriebs, was einen vielversprechenden Weg für zukünftige Entwicklungen auf diesem Gebiet darstellt.

Die Verbesserung der Datenfilterung anhand der Wetterdaten, insbesondere Sonneneinstrahlung und Wolkenbedeckung, ist eine Priorität, um die Genauigkeit und Geschwindigkeit der Analyse zu erhöhen. Darüber hinaus könnten Informationen über die Orientierung des Solarkollektors, wenn sie im Datensatz vorhanden sind, die Schätzung des Solarertrags in der RBA verbessern und können auch als exogenes Attribut in die MLA implementiert werden. Die nächsten Schritte umfassen eine eingehende Literaturrecherche zur Implementierung von Algorithmen des maschinellen Lernens bei der Fehlererkennung in Gebäudeenergiesystemen und deren Anpassung an solarthermische Systeme. Gleichzeitig soll die in diesem Projekt getestete MLA-Pipeline für eine nahtlose Integration in das bestehende Produktivsystem weiterentwickelt werden. Diese Bemühungen zielen gemeinsam darauf ab, die Zuverlässigkeit und Effizienz solarthermischer Systeme in praktischen Anwendungen zu verbessern.

Mit Blick auf die Zukunft zielt die Forschung darauf ab, die Strategien zur Fehlererkennung auf Wärmepumpensysteme auszuweiten und dabei den Erfolg der regelbasierten und maschinellen Lernalgorithmen zu nutzen.



## 6 Literaturverzeichnis

1. Sitzmann, B. *Qualitätsprüfung thermische Solaranlagen 2013*; Energie Zukunft Schweiz: AUE Basel-Stadt und AUE Basel-Landschaft, 2013.
2. Sitzmann, B. *Qualitätssicherung Förderprogramm "Sonnenkollektoren für Warmwasser und Heizungsunterstützung"*; Kanton Solothurn, Amt für Wirtschaft und Arbeit, Energiefachstelle: 2014.
3. Sitzmann, B. *Qualitätsprüfung thermische Solaranlagen 2014, Schlussbericht*; Energie Zukunft Schweiz: AUE Basel-Stadt und AUE Basel-Landschaft, 2014.
4. Sitzmann, B. *Qualitätsprüfung thermische Solaranlagen 2015, Schlussbericht*; Energie Zukunft Schweiz, AUE Basel-Stadt und AUE Basel-Landschaft, Basel, 2015.
5. Sitzmann, B. *Qualitätsprüfung thermische Solaranlagen 2016, Schlussbericht*; Energie Zukunft Schweiz: AUE Basel-Stadt und AUE Basel-Landschaft, 2016.
6. Sitzmann, B. *Qualitätsprüfung thermische Solaranlagen 2018 Kantone Basel-Stadt und Basel-Landschaft - Schlussbericht*; Energie Zukunft Schweiz: AUE Basel-Stadt und AUE Basel-Landschaft, Basel, 2018.
7. Sitzmann, B. *Qualitätsprüfung thermische Solaranlagen 2019 Kanton Basel-Stadt - Schlussbericht*; Energie Zukunft Schweiz: AUE Basel-Stadt und AUE Basel-Landschaft, Basel, 2019.
8. de Keizer, A.; Vajen, K.; Jordan, U. 376-Overview of Monitoring and Failure Detection Approaches for Solar Thermal Systems. 2008.
9. Hofmann, J.W.; Sitzmann, B.; Dickinson, J.; Kunz, D.; Eismann, R. *LoCoSOL – LowCost-Monitoring von thermischen Solaranlagen*; Bundesamt für Energie BFE, Bern, 2021.
10. Hofmann, J.W.; Sitzmann, B.; Dickinson, J.; Kunz, D.; Eismann, R. Use of Machine-Learning for Monitoring Solar Thermal Plants. In Proceedings of CISBAT 2021, EPFL, Lausanne, 8.-10. September.
11. Sigrist, D.; Kessler, S. Harmonisiertes Fördermodell der Kantone (HFM 2015) - Schlussbericht. Bundesamt für Energie, BFE: Bern, 2015.
12. Swissolar, S.F.f.S. QM-Solarwärme. Available online: <https://www.swissolar.ch/fuer-fachleute/hilfsmittel-solarwaerme/tools/qm-solarwaerme/> (accessed on 18-05-2021).
13. G. Faure, M. Vallée, C. Paulus, and T. Q. Tran, "Fault detection and diagnosis for large solar thermal systems: A review of fault types and applicable methods," *Solar Energy*, vol. 197, pp. 472–484, Feb. 2020, doi: 10.1016/j.solener.2020.01.027.
14. L. Feierl, V. Unterberger, C. Rossi, B. Gerardts, and M. Gaetani, "Fault detective: Automatic fault-detection for solar thermal systems based on artificial intelligence," *Solar Energy Advances*, vol. 3, p. 100033, 2023, doi: 10.1016/j.seja.2023.100033.
15. Josef Perktold, „statsmodels/statsmodels: Release 0.14.1“. Zenodo, Dez. 14, 2023. doi: 10.5281/zenodo.10378921.