

# Effektive Datenaufbereitung für prädiktive Modellierung mit Visplore

Harald Piringer  
piringer@vrvis.at



## VRVis: Kompetenz und Technologie

### Österreichs führendes Zentrum für Visual Computing

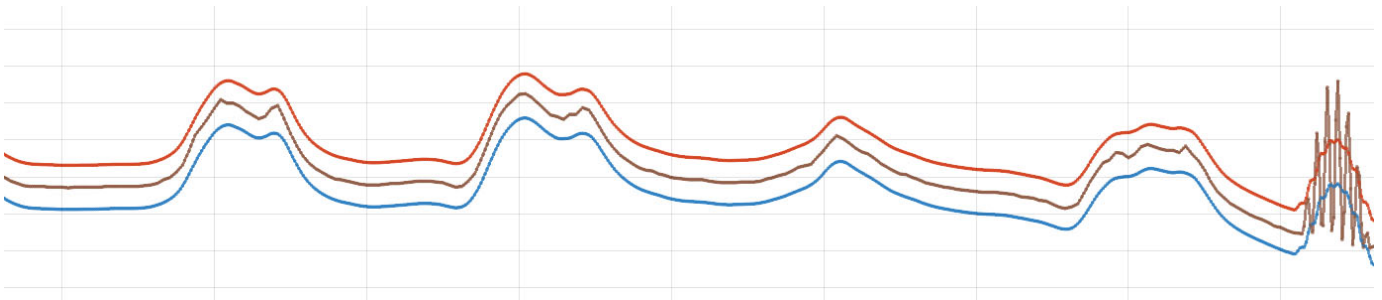
- Ca. 70 Mitarbeiter
- ARES-Tower in 1220 Wien
- Fokus: F+E und Data Science Projekte mit der Industrie
- COMET K1 Zentrum



# Ziel: Frühe und zuverlässige Erkennung von Störungen

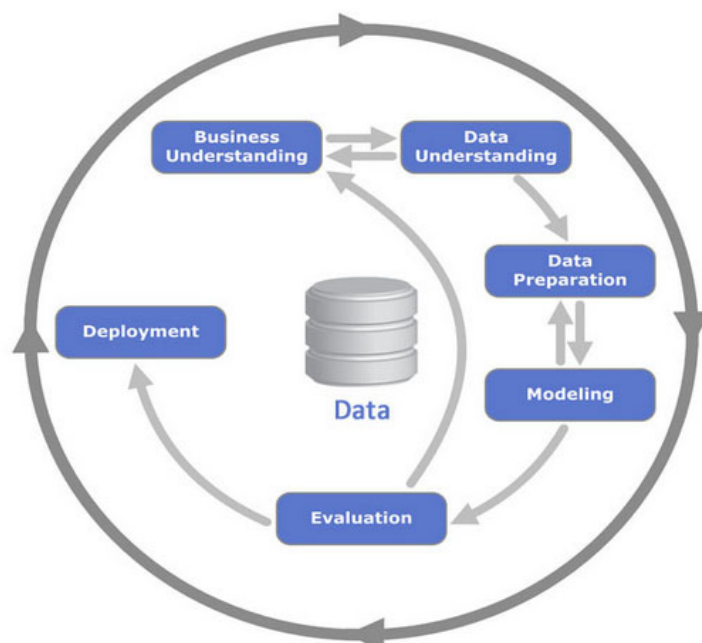
... erfordert meist eine prädiktive Datenmodellierung

- Dynamische Toleranzkorridore
- Abschätzungen der Restlebensdauer
- Zustandsklassifikation
- Etc.



## Prädiktive Modellierung ist ein aufwändiger Prozess

### CRISP-DM Process Diagram



Source: Kenneth Jensen

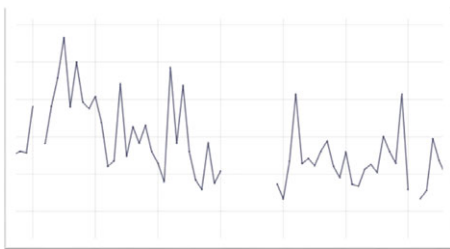
# Der „Daten“-Teil macht 80% des Aufwands aus

## CRISP-DM Process Diagram

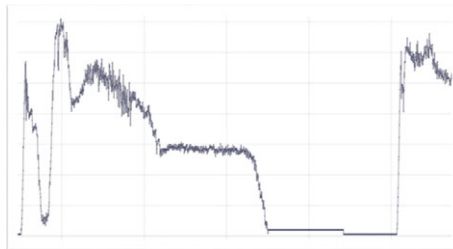


Source: Kenneth Jensen

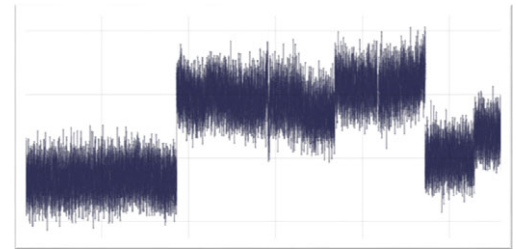
## Hürde 1: Datenqualität



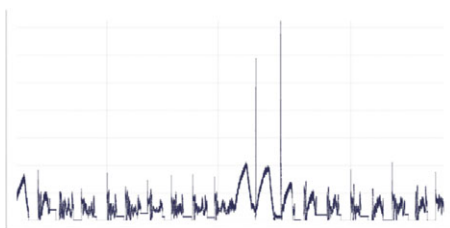
Fehlwerte



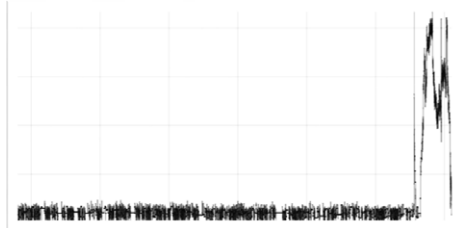
Wertwiederholungen



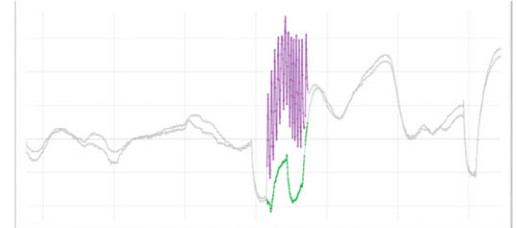
Offsets



Ausreißer



Strukturbrüche



Multivariate Anomalien

# Hürde 2: Einbringung von Prozesswissen

## Hürden für Digitalisierung

In welchen Bereichen bereitet die Digitalisierung Probleme?



Basis: Befragte, die Probleme bei der Digitalisierung haben (n = 66), Mehrfachnennung möglich, Mittelwerte, ©Cappgemini 2018

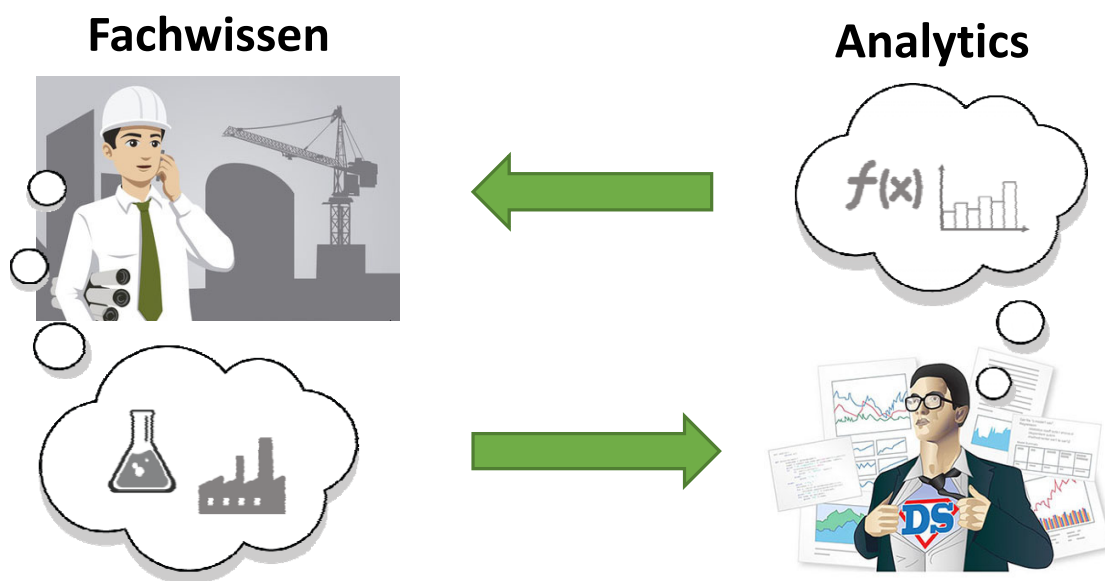
According to studies, companies lack 2-4 million "Business Translators" who master both specialist knowledge and analytics.

[McKinsey, „The Age of Analytics, 2016]

Reports stress that the need for data scientists can only be met if a large proportion of the existing domain experts are upskilled to help.

[McKinsey, „Skill Shift Automation and the Future of the Workforce“, 2018]

# Mission: Lücke zwischen Fachwissen und Analytik verkleinern



**Fachexperten ohne Data Science Hintergrund tiefgehende Analysen ermöglichen  
Zeit und Kosten fürs Verstehen, Überprüfen und Aufbereiten von Daten reduzieren**

# Ansatz: Interaktive Visualisierung

4) Dies führt zu weiteren Grafiken und Einsichten innerhalb von Sekunden

3) ... als Input für intelligente Algorithmen

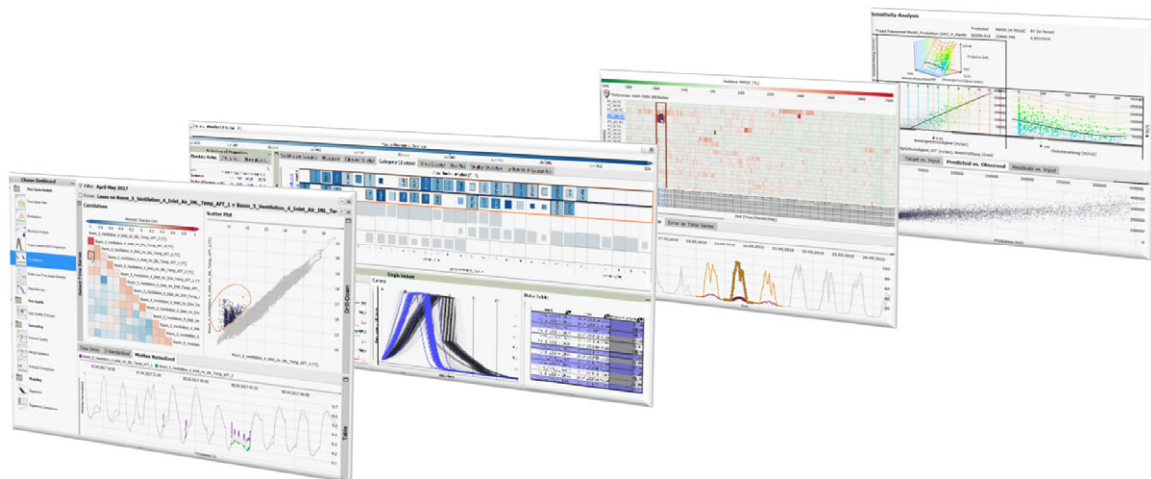


1) Benutzer gewinnen neue Einsichten aus Grafiken ...

2) ... wählen interessante Muster aus ...

## visplöre: Self-Service Analytik für Ingenieure

- **Vorkonfiguriert:** Sofortiger Zugang für Prozessexperten
- **Performant:** Mit Millionen Messungen in Sekundenbruchteilen interagieren
- **Vielseitig:** Qualitäts-, Prozess-, und Kontext-Daten korrelieren

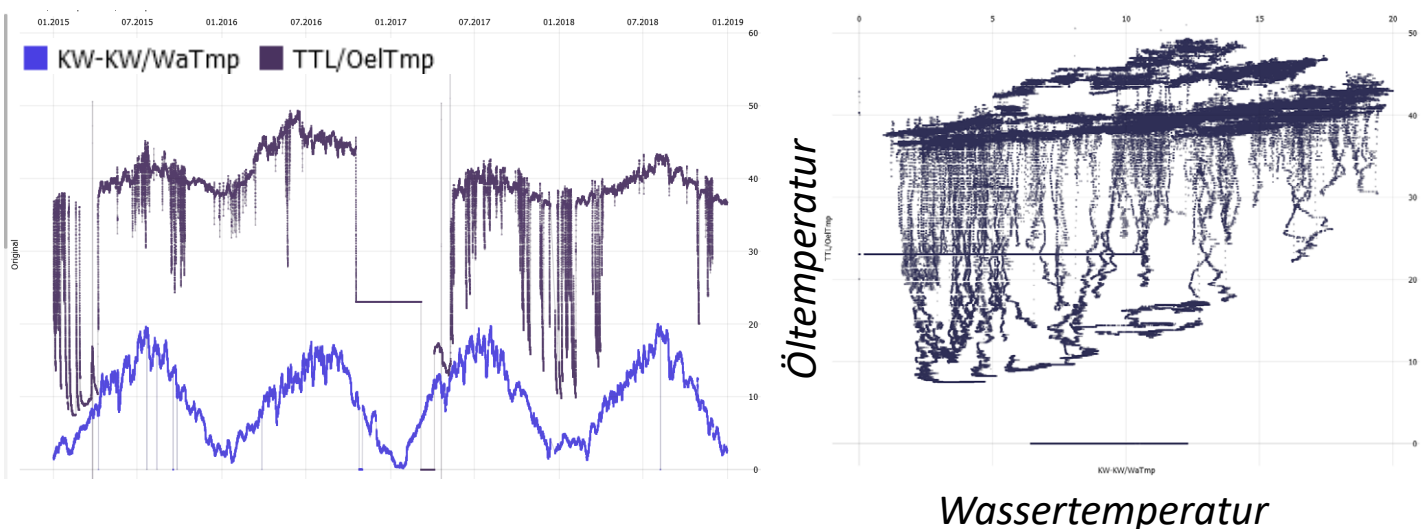


# Use Case: Modellierung von Turbinen-Prozessdaten

- **Ziel:** Frühzeitige Warnungen bei Sensoranomalien
- **Daten**
  - 21 Prozesszeitreihen (Temperaturen, Ölstände, etc.)
  - 4 Jahre in Minutenraster -> ca. 2 Mio. Datensätze
- **Ansatz:** Modellierung des Normverhaltens zur Berechnung dynamischer Toleranzkorridore
- **Zielgröße im Beispiel:** Öltemperatur des Turbinentraglagers ("TTL/OelTmp")
- **Einflussgrößen:** Wassertemperatur und andere Sensoren

## Herausforderung: Datenauswahl

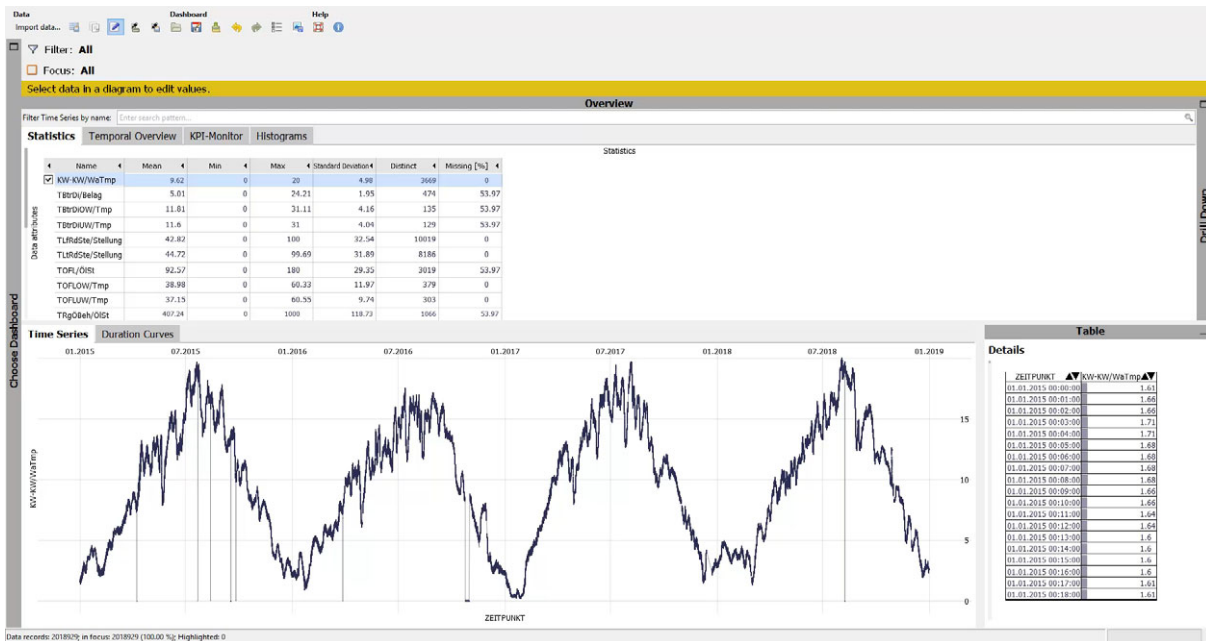
### Rohdaten von Öltemperatur und Wassertemperatur



**Direkte Modellierung nicht möglich -> Datenaufbereitung nötig**

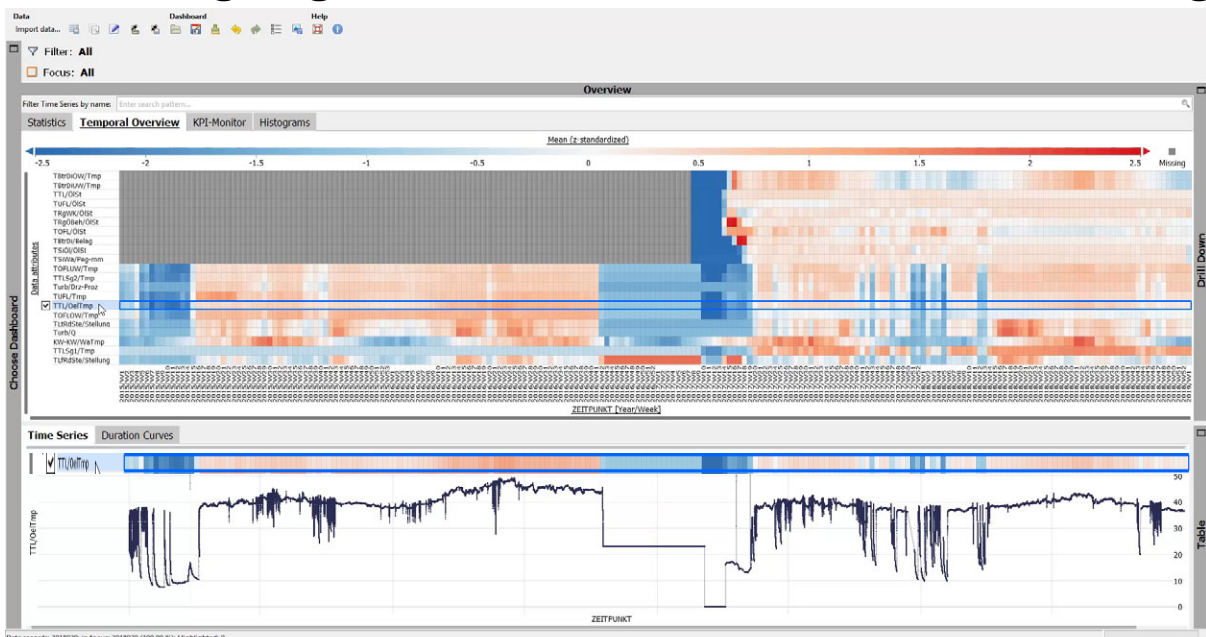
# Schritt 1: Bereinigung von Ausreißern

## Ersetzen von Ausreißern durch Interpolation



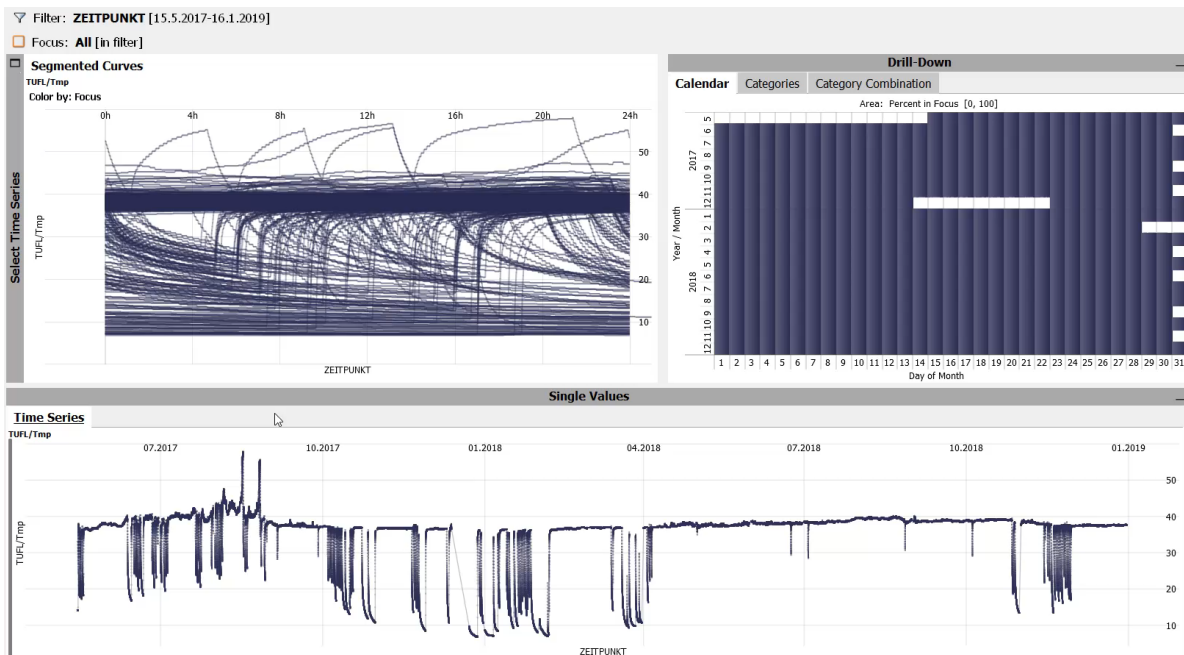
# Schritt 2: Datensichtung

## Identifikation geeigneter Zeiträume für eine Modellierung



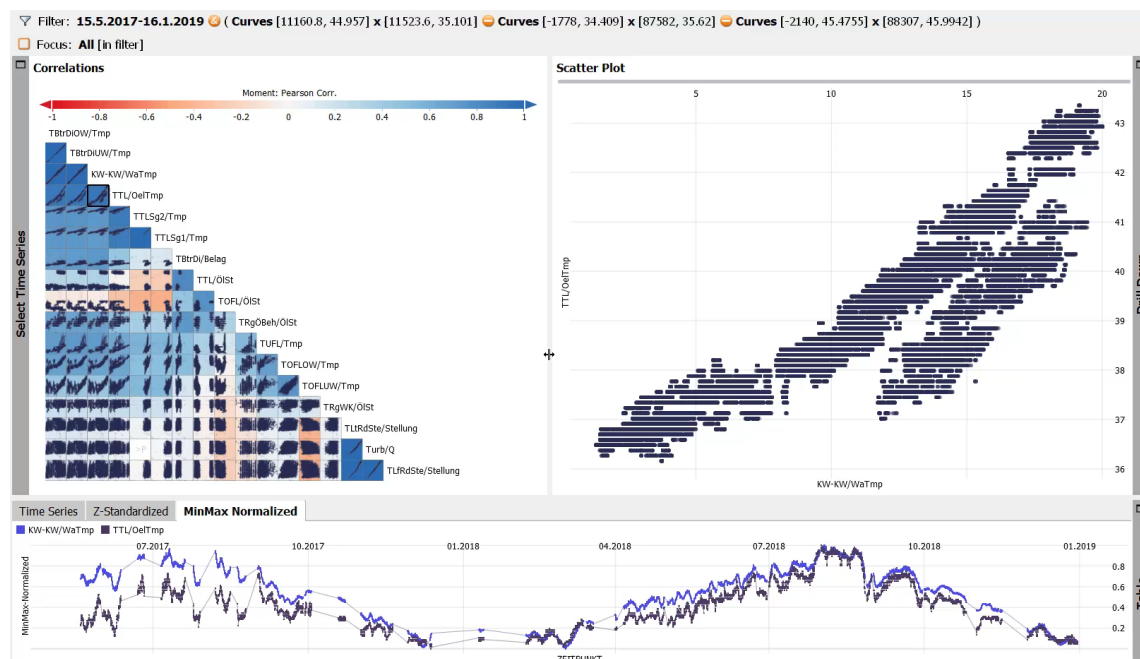
# Schritt 3: Einschränkung auf „Normalbetrieb“

Ausschluss von Tagen mit Abschalt- und Anfahrvorgängen



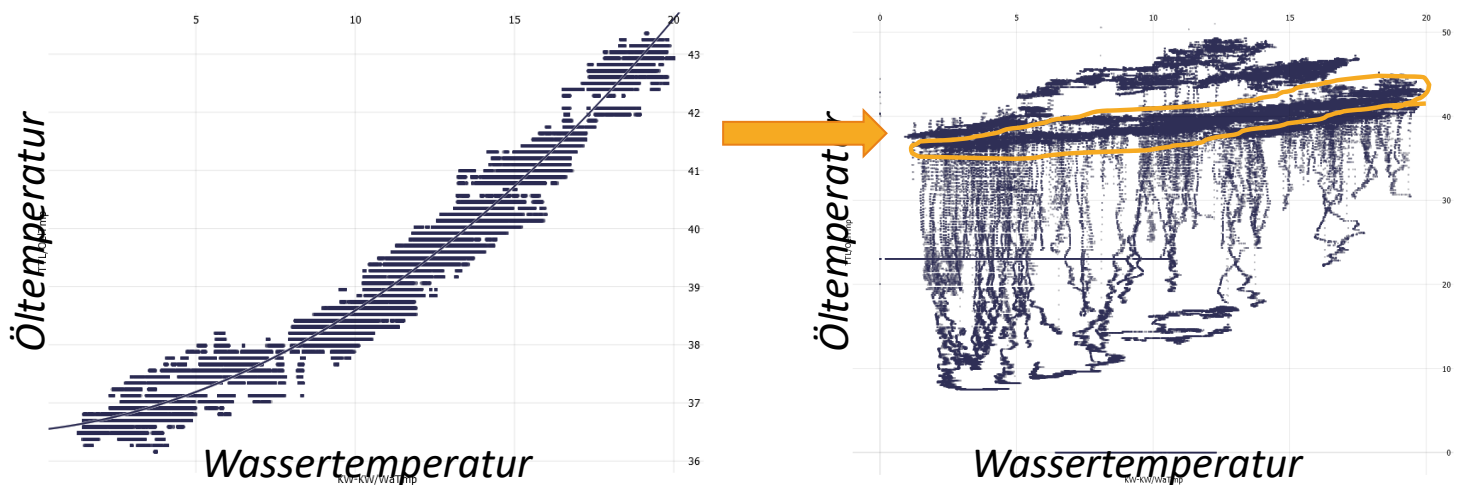
# Schritt 4: Analyse von Korrelationen und Clustern

Strukturbrüche durch Korrelationsanalyse aufspüren



## Schritt 5: Modellierung (schematisch)

Quadratisches Polynom liefert **jetzt** einen guten Fit

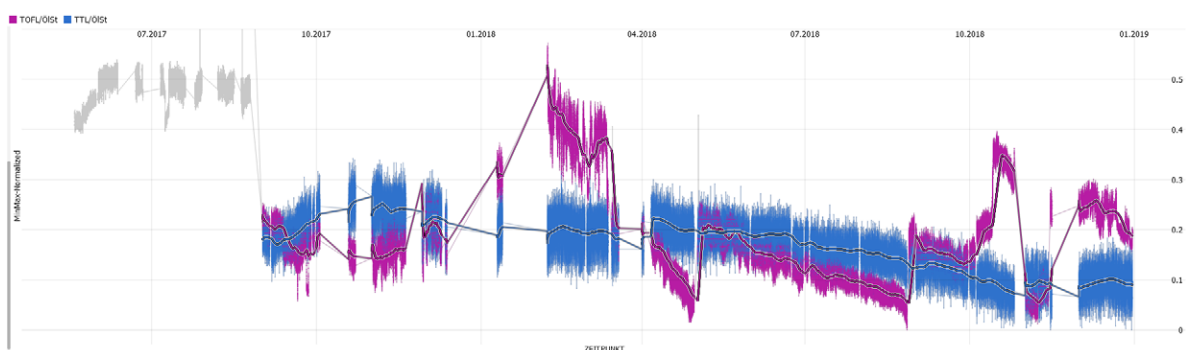


$$\text{Öltemp} = 0.015 x^2 + 0.058 x + 36.53$$

(x = Wassertemperatur)

## Ergebnisse

- Gutes UND einfaches Modell durch gezielte Datenauswahl
- Visuell und ohne eine einzige Zeile Programmcode
- Erkenntnisgewinn über die Daten als „Nebeneffekt“
- Viele weitere Entdeckungen wären einen Klick entfernt ...

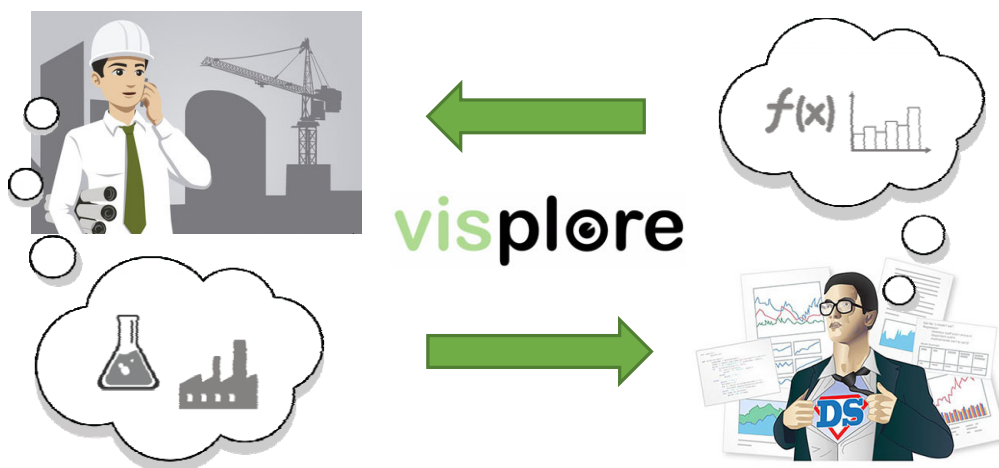


Z.B. Stufenmuster in den Ölständen ...

# Schlussfolgerungen

---

- Störungserkennung erfordert oft eine prädiktive Modellierung
- 80% des Aufwands resultiert aus Datenaufbereitung und der Kommunikation zwischen Data Scientists und Prozessexperten
- Interaktive Visualisierung ist effizient für die Validierung, Auswahl und Kommunikation UND für Nicht-Data Scientists leichter zugänglich als Programmiersprachen
- Visplore wurde entwickelt, um die Lücke zwischen Fachwissen und fortschrittlicher Analytik in der heutigen digitalisierten Industrie zu schließen



**Ich freue mich auf ein Gespräch mit Ihnen!**

---

piringer@vrvis.at  
www.vrvis.at

# Referenzen

---

## Industrieproduktion



## Energiewirtschaft



## Automobil- entwicklung

