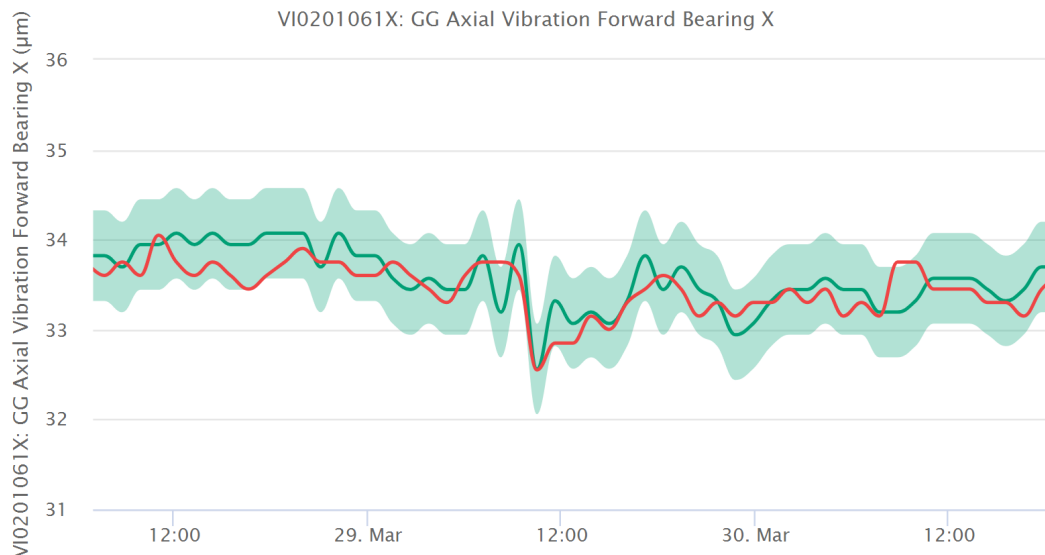
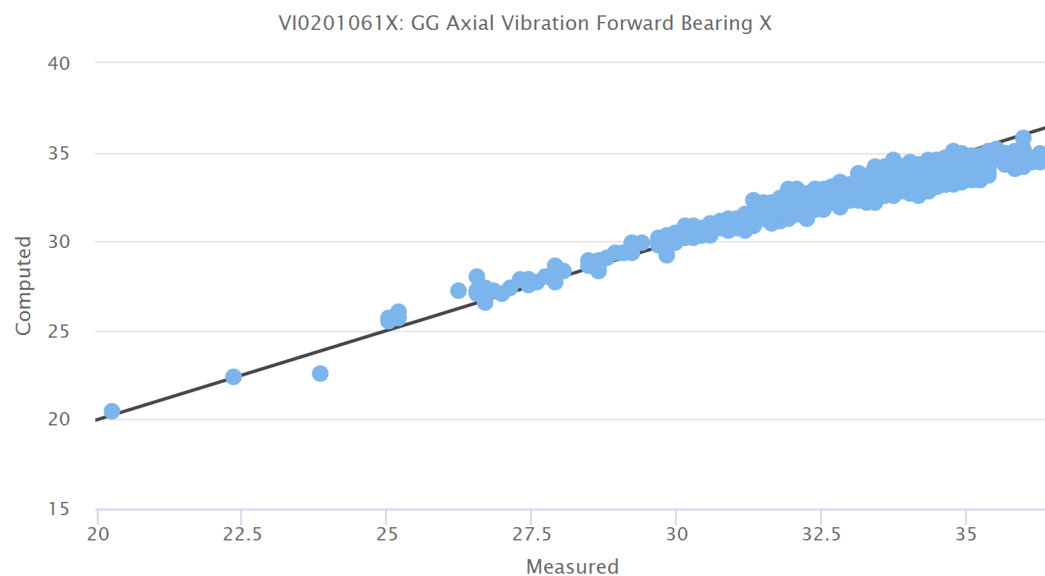


Handbuch: 7.1.5. Die Modellqualität beurteilen

Das Modell sollte die Trainingsdaten sehr gut reproduzieren. Der echte Test für die Eignung des Modells liegt allerdings bei jenen Daten, mit denen das Modell nicht trainiert wurde. Aus diesem Grund wird der von Ihnen für die Zeitspanne des Trainings spezifizierte Datensatz automatisch unterteilt in einen Datensatz fürs Training und einen Datensatz zur Validierung.

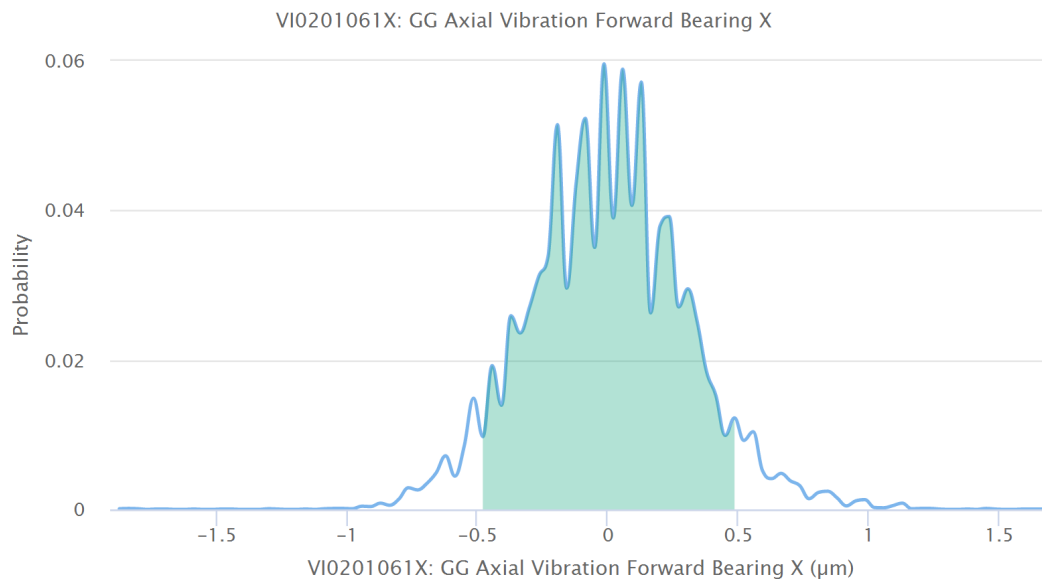


Nach dem Training wird das Modell automatisch auf seine Eignung taxiert, und das kann in der Oberfläche auf verschiedene Weise angeschaut werden. Der direkteste Weg ist die historische Grafikdarstellung von Modell und Messung, bei der Sie die zeitliche Entwicklung sowohl des Modells als auch der Messungen nachvollziehen können. Wenn beide während des optimalen Gesundheitszustandes nahe beieinander und bei suboptimalem Gesundheitszustand weit auseinander befinden, dann funktioniert das Modell gut.



Im Allgemeinen erwarten wir, dass es für die Abweichungen zwischen Modell und Messung kaum Unterschiede gibt, wenn wir den Datensatz fürs Training mit dem

Datensatz für die Validierung vergleichen, weil beide von dem optimalen Gesundheitszustand der Maschinerie ausgehen. Vergleichen wir in der grafischen Darstellung Modell mit Messung, so erwarten wir eine gerade Linie bei 45 Grad für beide Achsen welche im Diagramm durch eine schwarze Linie dargestellt wird.



Zusätzlich dazu erwarten wir auch Abweichungen zwischen Modell und Messung dergestalt, dass die meisten davon nur sehr klein sind. Stellen wir diese Abweichungen auf der horizontalen Achse dar und den Anteil der Punkte, die abweichen, auf der vertikalen Achse, so ergibt sich eine Wahrscheinlichkeitsverteilung der Abweichungen. Auch das wird in der Oberfläche dargestellt. Wir erwarten hierfür eine glockenförmige Kurve, ideal für eine normale oder Gaußsche Verteilung. Ist das der Fall, so ist dies ein weiterer Indikator für eine gute Modellqualität.

Wir messen auch die statistischen Größen wie Durchschnitt, Standardabweichung, Schiefe, und Wölbung für diese Wahrscheinlichkeitsverteilung, um weitere Details anzubieten. In der Praxis reicht eine visuelle Betrachtung dieser drei Grafiken aus, um zu beurteilen, ob das Modell gut ist oder nicht.

Ist das Modell nicht gut, gibt es dafür zwei mögliche Gründe. Der erste könnte sein, dass die von Ihnen ausgewählten unabhängigen Variablen nicht die richtigen sind. Entweder müssen wir dann einige weitere hinzufügen oder einige störende entfernen. In der Regel bedarf es der Hinzufügung von weiteren Variablen. Zweitens könnte es sein, dass die ausgewählten Trainingszeiten nicht lang genug waren oder auch ungesunde Gesundheitszustände enthielten. Bitte überprüfen Sie das sorgfältig, weil die Integrität der Daten, die verwendet werden, um einen gesunden Zustand zu definieren, den entscheidenden Faktor für die gesamte Analyse ausmachen.