

Report No. 4

Laden der CLIMAT-Temperaturdaten

Jürgen Grieser
(DWD/GPCC)

Von FE26 (Seewetteramt Hamburg) wurde dem GPCC ein Datensatz mit CLIMAT-Monatsmitteltemperaturen von 3155 Stationen bereitgestellt. Die Zeitreihen umfassen durchschnittlich fast 25 Jahre. Zusätzlich zu den Beobachtungswerten sind einige weitere Informationen zur Belegung mit täglichen Daten sowie der Herkunft der Daten angegeben.

Die Stationen sind durch ihre WMO-Nr. eindeutig charakterisiert. Sie sollen in das Messnetz 11 (WZN CLIMAT Niederschlag) geladen werden. Um das Laden durch das Lademodul zu ermöglichen, müssen zumindest die Koordinaten der zu ladenden Stationen angegeben werden. Für einen reibungsarmen Betrieb des Lademoduls ist es jedoch notwendig, alle Metainformationen in exakter Übereinstimmung mit den in der Datenbank bereits angelegten Stationen anzugeben.

Daher mussten diese Metainformationen zunächst aus der Datenbank extrahiert und den einzelnen Stationen zugeordnet werden. So wurde der ursprüngliche File mit 77.972 Zeilen, die jeweils die Informationen eines Stationsjahres tragen, ohne Hinzufügen neuer Information auf 1.856.304 Zeilen aufgebläht.

Für 64 der 3155 Stationen lag kein Eintrag der WMO-Nummer in der Datenbank-Tabelle Stationen_Kennungen vor. Diese Stationen scheinen Mirakel nicht unter der angegebenen WMO-Nummer bekannt zu sein. Für 17 der 64 Stationen konnte Udo Schneider herausfinden, dass ein Wechsel der WMO-Nummer stattgefunden hat. Er stellte die aktuelle, im GPCC verwendete, WMO-Nummer bereit. Für weitere 5 Stationen konnte er die notwendigen Metainformationen aus anderen Quellen bereitstellen. Eine Nachfrage beim Seewetteramt in Hamburg nach den Metainformationen der verbleibenden Stationen ergab für weitere 10 Stationen einen WMO-Nummerwechsel und für weitere 4 Stationen zusätzliche Metainformationen. Alle 27 Stationen mit WMO-Nummerwechsel konnten mit Hilfe der aktuellen WMO-Nummer einer Mirakelstation zugeordnet werden. Von den 9 Stationen, für die die Metainformationen aus anderen Quellen verfügbar waren, wurden 8 neu angelegt. Eine konnte einer bereits bestehenden Station zugeordnet werden.

Der Versuch, die Daten im dafür vorgesehenen Format zu laden, ist mehrfach gescheitert, da das Lademodul nicht die vorab behaupteten Eigenschaften aufwies. Es war aufgrund innerer

Widersprüche prinzipiell nicht in der Lage, Temperaturdaten zu laden. Nach mehrwöchigen Versuchen und Nachbesserungsarbeiten am Lademodul konnten die Daten von 3127 Stationen geladen werden.

Entsprechend den Vorgaben wurden Monatswerte mit bis zu 3 Fehltagen als 90% belegt und mit bis 7 Fehltagen als 80% belegt eingestuft. Monatswerte mit mehr als 7 Fehltagen wurden nicht in die Datenbank aufgenommen.

Die Temperaturdaten der 28 verbleibenden Stationen mit den folgenden WMO-Nummern wurden nicht geladen. Ihnen konnte keine Metainformation zugeordnet werden:

2087, 40454, 40476, 40490, 40530, 40575, 40586, 40598, 40620, 40640,
40661, 40675, 40715, 40756, 40846, 60469, 61802, 64118, 78763, 82572,
84045, 84073, 84129, 84162, 86576, 87161, 94986, 99031.

Folgende Schlussfolgerungen legt das Laden dieses Datensatzes nahe:

1. Die Entscheidung, mit der Belegung bei Temperatur genauso zu verfahren wie beim Niederschlag, ist keine leichte. Dies begründet, den langen Entscheidungsfindungsprozess von mehreren Monaten.
2. Eine Metadatenkontrolle zum Laden zwingend vorzuschreiben, ist unsinnig. Es verschwendet Arbeitszeit (in diesem Fall etwa 7 Wochen) ohne den geringsten Nutzen. Das Aufblähen der Datenfiles um einen Faktor 24 ist unnötig und unhandlich. Das Extrahieren von Metadaten aus der Datenbank und Zuordnen zu den einzelnen einzulesenden Daten ist unnötig fehleranfällig.
3. Das Lademodul war technisch nicht in der Lage, Monatswerte der Temperatur zu laden. Durch Versuch und Irrtum mussten die Gründe dafür erforscht werden. Nun lädt es zwar auch Monatstemperaturen, nicht aber ohne fälschlicherweise die Information auszugeben, dass es einige unklare Werte nicht geladen hat.