

Jakob Herz, Svenja Carrigan, Christoph Beecken, Oliver Kornadt

Bemessungsverfahren für den sommerlichen Wärmeschutz unter Verwendung von KI-Methoden

Der Klimawandel führt in Europa zu steigenden Oberflächentemperaturen, was besonders die gebaute Umgebung anfällig für Hitzeereignisse macht. Für die Aufgabe, den sommerlichen Wärmeschutz von Gebäuden zu evaluieren, wurde die Verwendung einer Methode des Maschinellen Lernens untersucht. Dieser Aufsatz beschreibt zunächst die Limitationen der Verfahren nach DIN 4108-2 und dann die Entwicklung eines Verfahrens, das auf dem Random Forest Regressor (RFR) basiert. Anschließend werden die Ergebnisse mit denen einer thermischen Gebäudesimulation verglichen. Der RFR wurde mit randomisierten Simulationsdaten trainiert, um die operative Temperatur und die Raumlufthtemperatur vorherzusagen. Die Evaluation zeigt, dass der RFR zwar qualitativ ähnliche Zeitreihen liefert, quantitativ sind jedoch signifikante Unterschiede festzustellen. Die Übertemperaturgradstunden werden im Vergleich zur Simulation systematisch unterschätzt, insbesondere bei extremen Wetterszenarien. Die Studie diskutiert die Herausforderungen ungleich verteilter Trainingsdaten und die Sensitivität der Ergebnisse gegenüber den verwendeten Testreferenzjahren. Der Ansatz bietet wertvolle Erkenntnisse, zeigt die Grenzen der gewählten Methode für die Bewertung des sommerlichen Wärmeschutzes auf und unterstreicht den Bedarf an weiteren Optimierungen.

Sommerlicher Wärmeschutz; Maschinelles Lernen; Random Forest Regressor; Thermische Gebäudesimulation; Übertemperaturgradstunden; DIN 4108-2

Evaluation Methods for Summer Thermal Protection Using AI Techniques

Climate change is causing rising surface temperatures in Europe, making the built environment particularly vulnerable to heat events. To evaluate the summer thermal protection of buildings, the use of a machine learning method was investigated. This paper first outlines the limitations of the methods specified in DIN 4108-2 and then describes the development of an approach based on the Random Forest Regressor (RFR). The results are subsequently compared with those of thermal building simulations. The RFR was trained using randomized simulation data to predict operative temperature and indoor air temperature. While the RFR produces qualitatively similar time series, significant quantitative differences are evident. Overtemperature degree hours are systematically underestimated compared to simulations, especially in

extreme weather scenarios. The study discusses the challenges of unevenly distributed training data and the sensitivity of results to the test reference years used. The chosen approach provides important insights while demonstrating the limitations of assessing summer thermal protection using the chosen method and highlights the need for further optimization.

summer thermal protection; machine learning; random forest regressor; thermal building simulation; overtemperature degree hours; DIN 4108-2

1 Einleitung

Der voranschreitende Klimawandel sorgt besonders in Europa für eine Erwärmung der Oberflächentemperaturen [1]. Vor allem die gebaute Umgebung ist anfällig gegenüber Hitzeereignissen [2], deren Häufigkeit über die letzten Jahrzehnte zugenommen hat [3] und deren weitere Zunahme prognostiziert wird [4]. Anpassungen an höhere Temperaturen im Gebäudesektor werden notwendig sein. Dabei ist es notwendig, den sommerlichen Wärmeschutz von neuen Gebäuden und Sanierungsobjekten bereits in der Planungsphase zu berücksichtigen und zu evaluieren. Besonderes Augenmerk verdient die heutige Berücksichtigung künftiger Einflüsse des Klimawandels, um zukunftssichere Gebäude zu erhalten. Die fortschreitende Entwicklung von Methoden der Künstlichen Intelligenz lässt vermuten, dass Methoden des Maschinellen Lernens (ML) verwendet werden können, um den sommerlichen Wärmeschutz von Gebäuden zu evaluieren. In diesem Aufsatz wird die Entwicklung eines auf ML basierten Verfahrens zur Bemessung des sommerlichen Wärmeschutzes beschrieben und die Ergebnisse mit denen einer thermischen Gebäudesimulation verglichen.

1.1 Mindestanforderungen an den sommerlichen Wärmeschutz nach DIN4108-2

Laut Gebäudeenergiegesetz [5] muss die Einhaltung der Mindestanforderungen an den sommerlichen Wärmeschutz [6] für Neubauten verpflichtend nachgewiesen werden. Dabei kann zwischen dem Verfahren der Sonneneintragskennwerte und der thermischen Gebäudesimulation gewählt werden.

Das Verfahren der Sonneneintragskennwerte ist ein tabellarisches Verfahren. Hierbei wird basierend auf empirischen Größen ein vorhandener Sonneneintragskennwert bestimmt, der von der Fläche der Fenster, ihrem Gesamtenergiedurchlassgrad und der Grundfläche des Raums abhängig ist. Dieser muss einen zulässigen Kennwert einhalten, der unter anderem die Nutzungsart, den Standort und Wärmeschutzmaßnahmen berücksichtigt. Das Verfahren ist zeitgünstig

durchzuführen, gilt jedoch als ungenau. Es beachtet zum Beispiel keine lokalen Umgebungsbedingungen. Weiterhin besteht keine Möglichkeit, zukünftige Klima- und Wetterprognosen mit einzubeziehen.

Bei dem Verfahren der thermischen Gebäudesimulation wird zunächst unter Vorgaben von Simulationsparametern die operative Temperatur (T_{OP}) im stündlich aufgelösten Jahresgang berechnet. Daraus werden die Übertemperaturgradstunden ($UETGS$) bestimmt. Für ein Wohngebäude werden diese nach

$$UETGS = \sum_{t=1}^{8760} \max(0, T_{OP,t} - T_{limit})$$

berechnet, wobei t die jeweilige Stunde im Jahr und T_{limit} die Grenztemperatur ist. Für die wärmste der drei Sommerklimaregionen, Region C, ist $T_{limit} = 27^\circ\text{C}$. Um den Nachweis zu erfüllen, darf ein Wohngebäude den Jahresgrenzwert von $UETGS = 1200 \text{ Kh}$ nicht überschreiten. Durch die Simulation können Einflüsse der lokalen Umgebung und der Gebäudegeometrie mitberücksichtigt werden. Prinzipiell lassen sich auch ortsaufgelöste oder prognostizierte Wetterdateien [7, 8] als Klimarandbedingungen der Simulation verwenden, laut DIN 4102-8 müssen jedoch die im Jahr 2011 erstellten Testreferenzjahre (TRJ) [9, 10] stellvertretend für die entsprechende Sommerklimaregion gewählt werden. Für Sommerklimaregion C ist dies das Normaljahr des Standorts Mannheim (TRJ2011_12). Die Problematik der Verwendung von veralteten Klimadaten wurde bereits diskutiert [11]. Während die Simulation für die gegebenen TRJ präzise Ergebnisse liefert, ist sie jedoch zeit- und arbeitsintensiv, sodass sie seltener Anwendung findet als das Verfahren der Sonneneintragskennwerte.

1.2 Zielsetzung

Im Rahmen eines Forschungsprojektes wurde eine Web-Applikation entwickelt, die bei einfacher Bedienbarkeit wenige Parameter eines Gebäudes bzw. Raums abfragt und basierend darauf eine Zeitreihe der Raumlufthtemperatur (T_{AIR}) und der operativen Temperatur T_{OP} berechnet, sodass eine Bewertung nach $UETGS$ möglich ist [12]. Die Vorhersage soll dabei vergleichbar der Genauigkeit einer Simulation von einem ML-Algorithmus erstellt werden. Dieser Aufsatz konzentriert sich auf das Training des ML-Algorithmus und dessen Vergleich mit einer Simulation.

2 Methodik

2.1 ML-Algorithmus: Random Forest Regressor

Die Vorhersage von T_{AIR} und von T_{OP} in einem Raum anhand von unabhängigen Eingangsvariablen (*features*) ist im Wesentlichen ein Regressionsproblem, bei dem eine Funktion den Zusammenhang zwischen den *features* und der abhängigen Ziel-Variablen (*target*) bestimmen soll. Die Regression kann durch verschiedene Methoden des überwachten ML erfolgen, etwa durch lineare oder polynomische Regressionsalgorithmen, aber auch die Verwendung neuronaler Netze ist denkbar. Der hier gewählte Ansatz ist der Random Forest Regressor (RFR) [13]. Er besitzt den Vorteil, dass er keine Vorannahme über eine zu parametrisierende Funktion trifft, wie es etwa bei der linearen Regression der Fall ist, so dass er nichtlineare Zusammenhänge abbilden kann. Weiterhin gilt er als robust gegenüber Überanpassung (*Overfitting*) [13] und die einzelnen Entscheidungsbäume (englisch: Decision Trees, DT), auf denen der RFR basiert, sind transparent und erklärbar. So kann die Wichtigkeit der einzelnen *features* festgestellt werden.

2.1.1 Ensemble aus randomisierten Decision Trees

Der RFR ist ein Ensemble-Algorithmus, der eine Vorhersage aufgrund einer Vielzahl von Vorhersagen einzelner DTs trifft. Ein einzelner trainierter DT teilt eine Datenmenge anhand von Entscheidungsknoten in mehrere Untermengen auf und weist ihnen einen Vorhersagewert zu, siehe Bild 1. Die Reihenfolge und der Wert der Entscheidungsknoten sowie die Werte der Untermengen werden während dem Training festgelegt. Durch *bagging* (*bootstrapping and aggregating*) werden beim Training eines RFRs viele verschiedene DTs erstellt. Dabei bedeutet *bootstrapping*, dass einem einzelnen DT nur eine zufällig gewählte Untermenge aus der gesamten Menge an Trainingsdaten zur Verfügung gestellt wird. Zudem wird eingeschränkt, anhand wie vieler *features* der jeweilige DT seine Entscheidungen treffen kann. *Aggregating* bezeichnet das Mitteln der Vorhersagen aller DTs bei der Abfrage des RFRs.

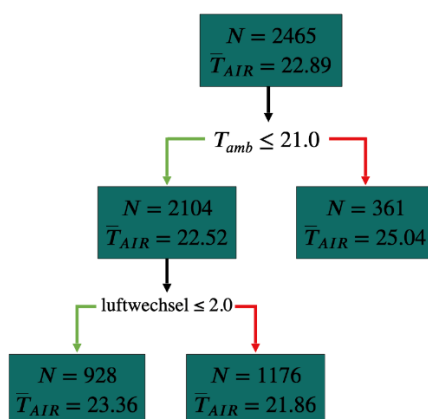


Bild 1 Ein einzelner Decision Tree. Ein Datenset mit N Datenpunkten wird anhand der Umgebungstemperatur T_{amb} und des

Luftwechsels in drei Datensets aufgeteilt.

2.1.2 Einschränkungen durch den RFR

Der verwendete RFR-Algorithmus [14, 15] ist darauf ausgelegt, tabellarische Daten zu verarbeiten. Bei der Abfrage eines trainierten RFR wird dem Algorithmus eine Menge an Datenpunkten übergeben, die er in Untermengen unterteilt. Ist diese Menge zu klein und umfasst bspw. nur einen einzelnen Datenpunkt, so wird die Rechenzeit durch den Overhead bestimmt. Daher wäre eine explizite Abfrage des RFRs, in der, wie in einer Simulation, ein Datenpunkt nach dem anderen abgefragt wird, nicht praktikabel. Da jedoch auch die thermische Historie eines Raums, also die Temperatur der vorangegangenen Stunde $T_{AIR}(t-1)$, die Innenraumtemperatur $T_{AIR}(t)$ zum Zeitpunkt t und ebenso die Luftwechselrate nach DIN 4108-2 beeinflusst, muss die Zeitreihe implizit bestimmt werden. Deshalb müssen bei der Abfrage für jeden Zeitpunkt Initialwerte für $T_{AIR}(t-1)$ und die Luftwechselrate eingesetzt werden. Über eine Vielzahl an Abfragen hinweg werden die zeitabhängigen Werte angepasst und T_{AIR} konvergiert.

Weiterhin ist der RFR in seiner Gültigkeit auf den Wertebereich der *features* und des *targets* in den Trainingsdaten beschränkt. So wurde vor der Generierung der Daten eine Liste an voraussichtlich notwendigen *features* und deren Wertebereich definiert, siehe Tabelle 1. Zudem muss eine Unabhängigkeit der *features* voneinander gewährleistet werden. Damit ergibt sich bei der Erstellung der Trainingsdaten die Notwendigkeit zur Randomisierung.

Tabelle 1: Übersicht über die gewählten *features* für die Abfrage des ML-Models.

feature	Beschreibung	min.	max.	Einheit
Raumeigenschaften				
länge	Kubatur	3,0	10,0	m
breite		2,5	6,0	m
höhe		2,4	4,0	m
etage	0: Erdgeschoss, 1: Mittelgeschoss, 2: Dachgeschoss	1	3	-
u_wert_wand	Wärmedurchgangskoeffizient der Außenwand, des Dachs, der	0,1	3,0	Wm ⁻² K ⁻¹
u_wert_dach _a	Bodenplatte und der Innenwand	0,1	3,0	Wm ⁻² K ⁻¹
u_wert_boden _a		0,1	3,0	Wm ⁻² K ⁻¹
u_wert_innen _b		0,5	3,0	Wm ⁻² K ⁻¹
luftwechsel		0,25	5,0	h ⁻¹
interne_lasten	Wärmeeintrag durch interne Quellen	0	50	kJh ⁻¹ m ⁻²
bauart	Thermische Kapazität pro Grundfläche	0	150	Wm ⁻² K ⁻¹
Fenstereigenschaften (für jedes Fenster)				

<i>fenster_zu_aussenflaeche</i>	Verhältnis von Fenster- zur Wandfläche	0,001	0,97	-
<i>rahmenanteil_fenster</i>	Verhältnis von Rahmen- zur Fensterfläche	0,01	0,4	-
<i>u_wert_fenster</i>	Wärmedurchgangskoeffizient des Fensters (inkl. Rahmen)	0,7	5,6	Wm ⁻² K ⁻¹
<i>g_wert_fenster</i>	Solarer Gesamtergiedurchlassgrad	0,14	0,84	-
<i>fc_wert_fenster</i>	Verschattungsgrad der äußeren Verschattung	0	1	-
Umgebungsparameter				
<i>RADOS_c</i>	Totale Sonnenbestrahlungsstärke auf vertikale Flächen	0	1000	Wm ⁻²
<i>T_{AMB}</i>	Außenlufttemperatur	-20	37,2	°C
<i>bewölkungsgrad</i>	Bewölkungsgrad	0	8	1/8

- a) Wird auch angenommen, falls der Raum sich im Mittelgeschoss befindet.
b) Existiert nicht als *feature* im ML-Model, wird aber in der Simulation berücksichtigt.
c) Wird aus der Horizontalstrahlung und der Orientierung des entsprechenden Fensters berechnet.

2.2 Generation von Trainingsdaten durch randomisierte Simulationen

Die Trainingsdaten für den RFR wurden durch Simulationen in TRNSYS18 generiert. TRNSYS18 ist ein etabliertes Simulationsprogramm für thermische Gebäudesimulationen [16]. Zunächst wurde ein Multizonen-Gebäude aufgebaut (Bild 2a), das sowohl 12 Eckräume mit je zwei Fensterfassaden als auch 12 Mittelräume mit je einer Fensterfassade besitzt.

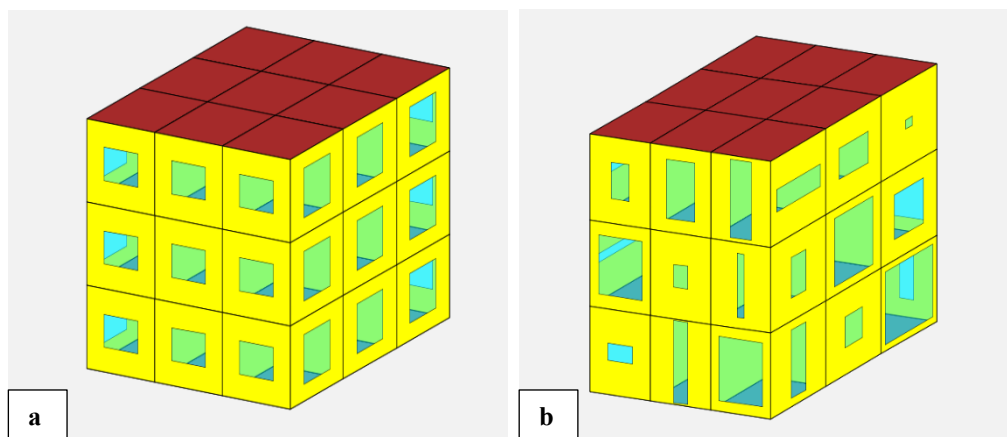


Bild 2 TRNSYS Multizonen-Gebäude, das zur Erzeugung von Trainingsdaten benutzt wurde. a) Referenzgebäude, um die automatisierte Simulation zu initiieren. b) Gebäudemodell nach der Randomisierung [17].

Um die notwendigen randomisierten Simulationen zu ermöglichen, wurde die Gebäudeinformationsdatei in TRNSYS mit in Python geschriebenen Algorithmen angepasst. Wie in [17] beschrieben, wurden dabei Funktionen entwickelt, die einerseits die für den RFR definierten *features* (Tabelle 1) aus den Simulationsdateien extrahieren und andererseits die Gebäude- und Zonenparameter so konfigurieren, dass die *features* möglichst gleichmäßig den für sie festgelegten Wertebereich abdecken. Die randomisierte Geometrie des Gebäudes ist Bild 2b zu entnehmen. Aufgrund der hohen Anzahl der *features* war es nicht praktikabel, für jeden

Simulationsdurchgang nur eine einzige Kombination an *features* zu simulieren. Deshalb wurde jeweils nach drei Simulationsstunden ein neuer Wert für den Luftwechsel, die internen Lasten, den Verschattungsgrad sowie den Rahmenanteil des Fensters angenommen.

Für jeden Simulationsdurchgang wurde eine neue Wetterdatei aus dem Klimadatensatz der TRJ 2017 des DWD [7, 8] ausgewählt. Dabei wurden die winter-fokussierten TRJs exkludiert und die sommer-fokussierten mit einer 75-prozentigen Wahrscheinlichkeit vor den Normaljahren bevorzugt. Es wurden 200 Simulationen durchgeführt, was ca. $41 \cdot 10^6$ Datenpunkte ergab.

2.3 Aufbereitung ungleich verteilter Daten

Thermische Gebäudesimulationen, die ein ganzes Jahr mit 8.760 Stunden abbilden und bei denen eine minimale Raumlufttemperatur T_{AIR} 20 °C, ab der die Heizung aktiviert wird, nicht unterschritten wird, erzeugen eine ungleiche Verteilung von T_{AIR} und T_{OP} . Während der Heizperiode akkumulieren sich Datenpunkte mit $T_{OP} \leq 20^\circ\text{C}$, während höhere Temperaturen Ausnahmereischeinungen bleiben, siehe Bild 3 links. Dies ist eine zu erwartende Temperaturverteilung für zentraleuropäisches Wetter.

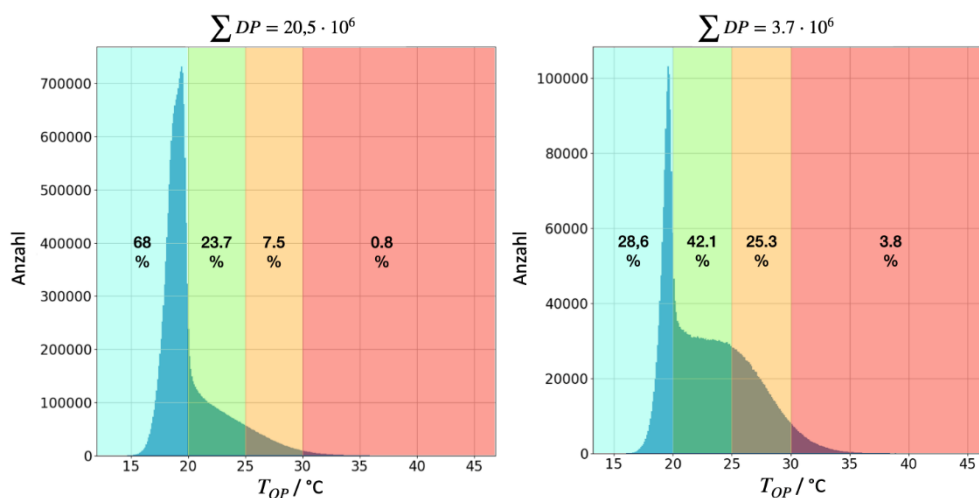


Bild 3 Verteilung der operativen Temperatur in den Trainingsdaten nach der Simulation (links) und nach der Aufbereitung der Daten (rechts).

Ungleich verteilte Trainingsdaten führen jedoch zu einem verzerrten RFR, der die unterrepräsentierten Szenarien nur ungenau vorhersagen kann. Um diesem Problem entgegenzuwirken, wurden verschiedene Methoden zum Datenausgleich angewendet:

- Bei den Simulationen wurden sommer-fokussierte Wetterdateien präferiert. Dies erhöhte den Anteil der Datenpunkte mit hohen Temperaturen.

- Für die Auswahl an Trainingsdaten wurden nur Räume inkludiert, die über 675 Überetemperaturgradstunden aufweisen, siehe Bild 4 links. Als Grenztemperatur wurden dabei 26 °C gewählt.
- Die Wahl der einzelnen Datenpunkte erfolgte durch einen Gauß-Filter ($\mu = 5000$, $\sigma = 1300$) über die Temperatur-Zeit-Verteilung, sodass Daten aus den Sommermonaten mit höherer Wahrscheinlichkeit gewählt werden als Daten aus den Wintermonaten, siehe Bild 4 rechts.

Ein Versuch, mehr Datenpunkte mit hohen Temperaturen durch einen Augmentation-Algorithmus zu erstellen (KNNOR-Reg-Algorithmus [18]), war nicht erfolgreich. Daher wurde allein mit den oben genannten Methoden ein neuer Trainingsdatensatz erstellt, dessen Temperaturverteilung in Bild 3 rechts zu sehen ist.

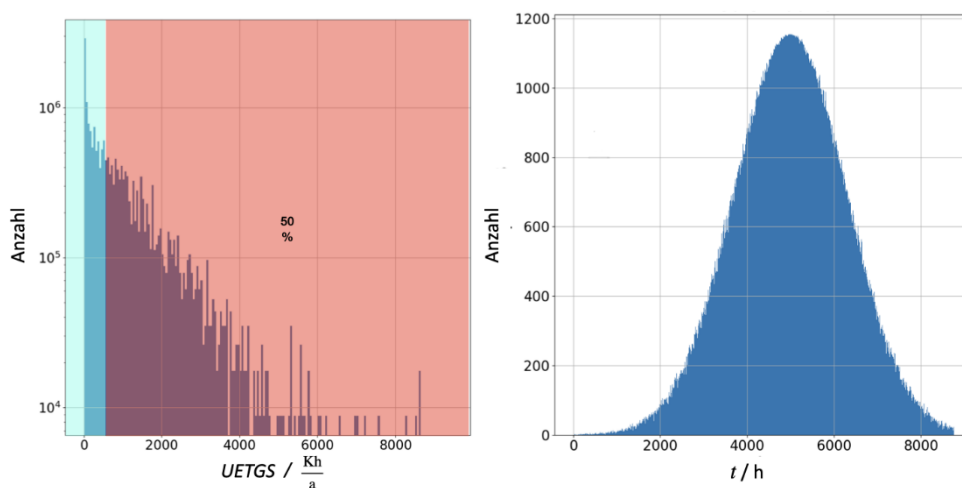


Bild 4 Strategien der Datenbereinigung: Wahl nach $UETGS$ links, Wahl nach Sommerstunden rechts.

2.4 Training von Random Forest Regressoren

Ein RFR kann nur ein einzelnes *target* vorhersagen. Da jedoch T_{AIR} und T_{OP} bestimmt werden sollen, müssen zwei verschiedene Regressoren trainiert werden. Dabei wird die in 2.1.2 erwähnte Historie nur für den T_{AIR} -Regressor implementiert, indem das *feature* $T_{AIR}(t - 1)$ mit aufgenommen wird. Der T_{OP} -Regressor hingegen nimmt T_{AIR} als *feature* auf, um die Rechenzeit-aufwendige implizite Abfrage zu vermeiden. Bei der Abfrage wird also zunächst T_{AIR} bestimmt und danach T_{OP} .

Weiterhin bedeutet die unterschiedliche Anzahl an Fenstern im betrachteten Raum eine unterschiedliche Anzahl an *features*, die das Modell zu beurteilen hat. Daher wurde für einen Eckraum und für einen Mittelraum jeweils ein Regressor für die Raumlufttemperatur und die operative Temperatur trainiert, sodass insgesamt vier Regressoren erzeugt wurden

(*RFR_TAIR_ecke*, *RFR_TAIR_mitte*, *RFR_TOP_ecke*, *RFR_TOP_mitte*).

Vor jedem Training wurden die Daten im Verhältnis 80 zu 20 in Trainingsdaten und Testdaten unterteilt, um die Regressoren nach dem Training anhand der Testdaten beurteilen zu können.

Als Bewertungskriterium wurde der mittlere quadratische Fehler

$mse = \sum_{i=1}^N \frac{(y_i - \hat{y}_i)^2}{N}$ gewählt. Nach einer Feinabstimmung zur Minimierung des *mse* wurden

die vier Regressoren mit folgenden Hyperparametern trainiert:

- Anzahl der einzelnen Entscheidungsbäume: 100,
- Anteil an betrachteten *features* pro Baum: 0,33,
- Relative Größe eines Datensets, das ein einzelner Baum betrachten kann, im Vergleich zur Größe des gesamten Trainingsdatensets: 0,4,
- Anzahl an Entscheidungen, die in einem Entscheidungspfad getroffen werden dürfen: keine Einschränkung,
- Minimalgröße einer Untermenge, sodass sie nicht weiter geteilt werden darf: 10.

3 Evaluation

Die Bewertung anhand des *mse* dient ausschließlich der Feinabstimmung der Hyperparameter.

Um die Genauigkeit der RFRs zur Vorhersage von Zeitreihen zu beurteilen, ist der *mse* nicht ausreichend. Die Evaluation erfolgt deshalb über einen Vergleich zwischen Zeitreihen, die durch eine thermische Gebäudesimulation erzeugt wurden, und Zeitreihen, die RFRs ausgeben.

Die Betrachtungen beschränken sich hier auf den Vergleich der operativen Temperaturen, da diese zur normativen Evaluierung des sommerlichen Wärmeschutzes verwendet werden, siehe 1.1.

3.1 Vergleich der Zeitreihen

Bild 5 zeigt den Jahresverlauf von *T_{OP}* als Ergebnis einer Simulation (links) und als Ergebnis des RFRs (rechts). Qualitativ weisen beiden einen ähnlichen Verlauf auf, im RFR-Verlauf sind jedoch deutlich nadelartige Artefakte zu erkennen. Weiterhin unterscheiden sie sich in ihren Maxima. Das Maximum bei Stunde 5.100 überschreitet bei der Simulation deutlich den Wert 30 °C, das Maximum bei Stunde 5.600 erreicht ungefähr einen Wert von 29 °C. Beide Maxima liegen bei dem RFR-Verlauf hingegen bei ca. 30 °C.

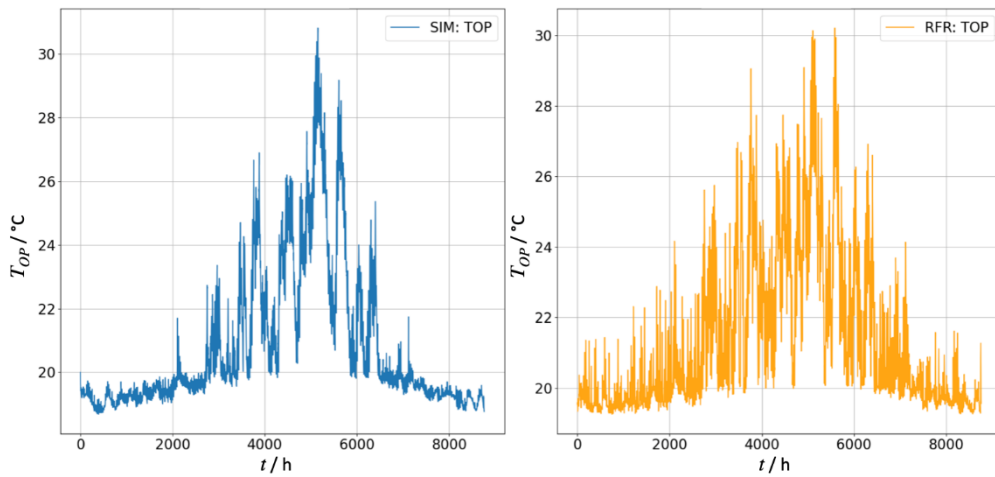


Bild 5 T_{op} -Zeitverlauf: links erzeugt durch eine thermische Gebäudesimulation, rechts erzeugt durch RFR-Verfahren.

Bei Betrachtung aller Werte, die in der Simulation oder dem RFR 25 °C überschreiten und für die Beurteilung des Wärmeschutzes bedeutend sind, siehe Bild 6 links, zeigt sich, dass der RFR die Simulationsergebnisse vor allem in der Übergangszeit überschätzt, während er sie in den Sommermonaten eher unterschätzt. Im Mittel schätzt der RFR die operativen Temperaturen etwas höher ein als die Simulation, wobei die Standardabweichung mit etwa 1,8 °C sehr hoch ist, siehe Bild 6 rechts.

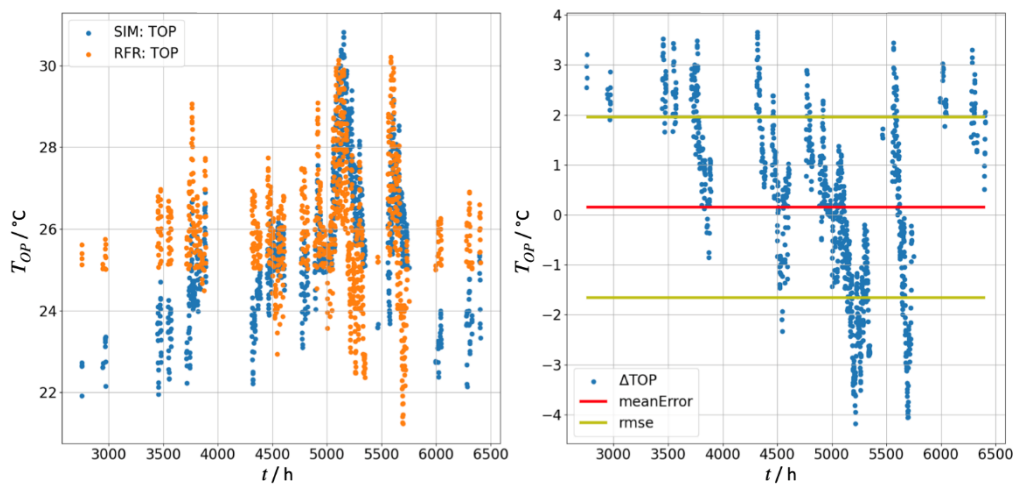


Bild 6 Links: Vergleich der operativen Temperatur oberhalb von 25 °C zwischen Simulation (blau) und dem RFR TOP_ecke (orange). Rechts: Differenz der Übertemperaturgradstunden (RFR – SIM, blau) zusammen mit dem mittleren Fehler (meanError, rote Linie) und der Standardabweichung (rmse, grüne Linien).

3.2 Bewertung nach Übertemperaturgradstunden

Die in 3.1 aufgeführte Evaluierung zeigt, dass der RFR die operative Temperatur – anders als die Simulation – nicht akkurat als Zeitreihe darstellen kann. Es stellt sich die Frage, ob die

Übertemperaturgradstunden des Raums trotz der Ungenauigkeiten in der Zeitreihe korrekt vorhergesagt werden können. Dazu wurde derselbe Raum anhand 400 verschiedener Wetterdateien für Mannheim getestet. Dabei wurde ein Quadratkilometer-Raster über eine Fläche von 100 km² gelegt, siehe Bild 7 links. Für die anschließende Simulation beziehungsweise RFR-Abfrage wurden sodann die Testreferenzjahre für das Normaljahr und das sommer-fokussierte Jahr der Jahre 2015 und 2045 als Wetterdatei zugrundegelegt. Anschließend wurden aus der errechneten T_{OP} -Zeitreihe die Übertemperaturgradstunden bestimmt. In Bild 7 rechts sind die Ergebnisse des RFRs gegen die Referenzwerte der Simulation aufgetragen. Bemerkenswert ist die lineare Korrelation zwischen den RFR-Ergebnissen und den Simulationsergebnissen. Die Steigung der Korrelationsgeraden ist jedoch bedeutend geringer als die der Winkelhalbierenden. Die Übertemperaturgradstunden werden also für die extrem heißen Wetterszenarien durch den RFR kategorisch unterschätzt. Beachtenswert ist, dass die sommerfokussierten TRJs 2015 höhere UETGS erzeugen als die prognostizierten Normaljahre 2045.

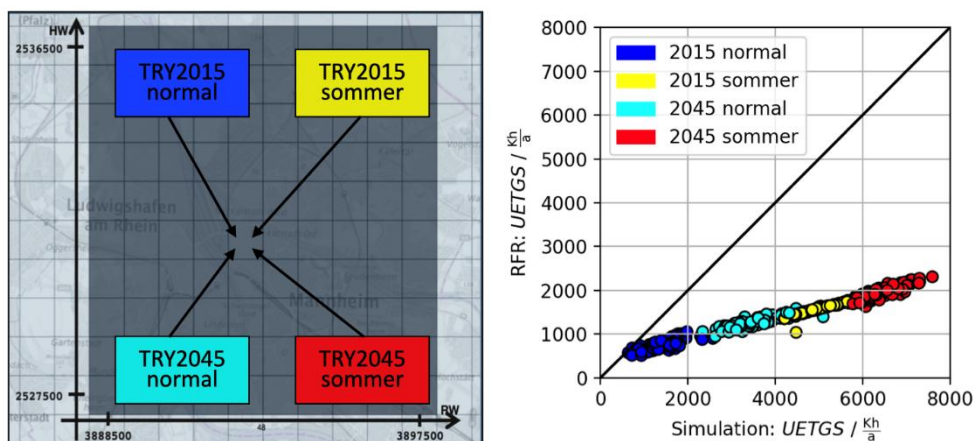


Bild 7 Links: 100 km² Raster über Mannheim, für jeden Quadratkilometer wurden 4 TRJ ausgewählt: das Normaljahr 2015, das sommerfokussierte Jahr 2015, das Normaljahr 2045 und das sommerfokussierte Jahr 2045. Rechts: Ermittelte UETGS ermittelt durch das RFR-Verfahren aufgetragen gegen simulierte UETGS.

3.3 Simulationsergebnisse im Kontext der DIN4108-2

In Bild 8 sind die UETGS des Simulationsverfahrens abgebildet. Zusätzlich wurde der Grenzwert für Wohngebäude (1.200 Kh/a) und das Ergebnis einer Simulation mit dem in DIN 4108-2 festgelegtem TRJ2011_12 aufgetragen. Zu erkennen ist, dass nur etwa ein Viertel der Simulationen mit dem Normaljahr 2015 als Umgebungswetter den Grenzwert unterschreiten. Die Simulation mit TRJ2011_12 resultiert in 737,8 Kh/a und bewertet damit den Wärmeschutz des Raums deutlich besser als die meisten anderen Simulationen.

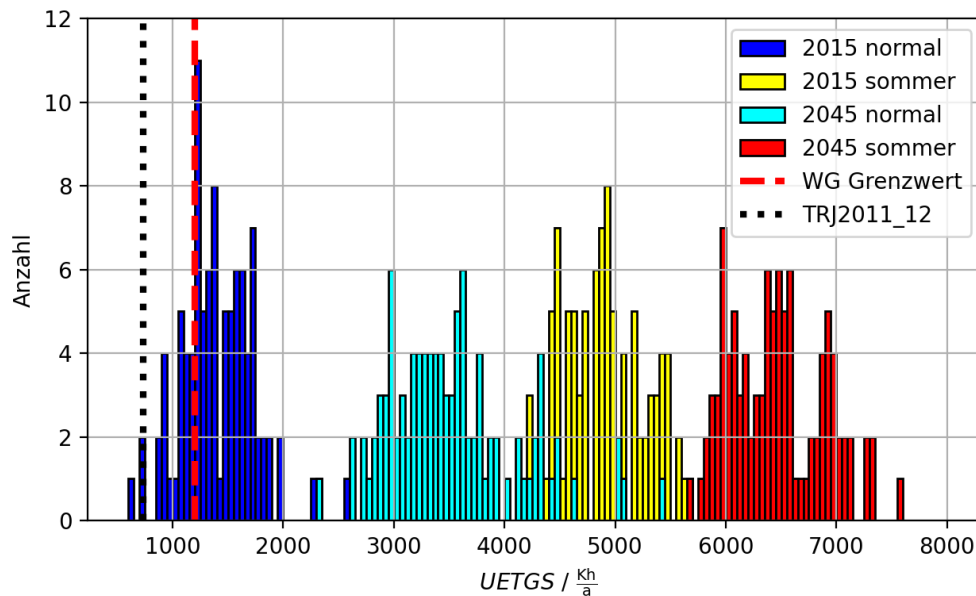


Bild 8 Verteilung der simulierten *UETGS* in alleiniger Abhängigkeit von den gewählten TRJs. Die gestrichelten Linien zeigen den Grenzwert für Wohngebäude (rot) und die *UETGS* für die entsprechende Simulation mit TRJ2011_12 (schwarz).

4 Diskussion und Ausblick

Der durch den RFR erstellte Zeitverlauf weist große quantitative Unterschiede zur simulierten Referenz auf. Um dem Phänomen der nadelartigen Artefakte entgegenzuwirken, könnte möglicherweise die Berücksichtigung einer längeren Historie über n Stunden mit $T_{AIR}(t - n)$ oder das Hinzufügen von Kontrollkriterien für die implizite Abfrage hilfreich sein.

Weiterhin wächst die Differenz in den *UETGS* zwischen Simulation und RFR umso mehr, je mehr *UETGS* die Simulation vorhersagt. Das RFR-Verfahren ist damit ungeeignet, um heiße Extremwetter zu evaluieren. Der Hauptgrund wird in der ungleichen Verteilung der Trainingsdaten vermutet. Hohe Temperaturen sind eine Ausnahmeerscheinung, wodurch sie beim Training unterrepräsentiert sind und fehlerhafte Vorhersagen zur Folge haben. Dies stellt eine signifikante Herausforderung für statistische Algorithmen dar, wenn das Interesse auf eben diesen Ausnahmeerscheinungen liegt. Zu beachten ist zunächst, dass Werte, die außerhalb der Trainingsdaten liegen, mit dem RFR-Verfahren grundsätzlich nicht vorhergesagt werden können.

Darüber hinaus bemerkenswert ist die Verteilung der simulierten *UETGS*. Sie zeigt eine hohe Sensitivität der *UETGS* gegenüber den gewählten TRJ. Eine Analyse der Wichtigkeit der *features* für die RFR-Vorhersage bestätigte dies. Eine Reduktion der *features* der Raum- und Flächeneigenschaften könnte daher sinnvoll sein. Dies wirft jedoch die prinzipielle Frage auf, welche Wetterdatei für eine Evaluation des sommerlichen Wärmeschutzes verwendet werden sollte.

Das hier vorgestellte RFR-Verfahren zur Vorhersage von T_{AIR} und T_{OP} ist nicht in der Lage, die Ergebnisse einer Simulation mit ausreichender Genauigkeit vorherzusagen und konnte nicht in der geplanten Web-Applikation verwendet werden.

Da jedoch die Bedienung der Simulationssoftware und nicht die (bei der RFR-Vorhersage geringere) Rechenzeit zeitkritisch für das simulative Verfahren ist, liegt der Fokus derzeit auf der Kopplung des Frontends der bedienungsfreundlichen Applikation mit einer leistungsfähigen thermischen Gebäudesimulation.

Danksagung

Die Autor:innen bedanken sich für die Förderung des Bundesministeriums für Wirtschaft und Energie im Zentralen Innovationsprogramm Mittelstand (ZIM).

Literatur

- [1] Schumacher, D.L. et al. (2024) *Exacerbated summer European warming not captured by climate models neglecting long-term aerosol changes*. Communications Earth & Environment 5, H. 1, S. 182. <https://doi.org/10.1038/s43247-024-01332-8>
- [2] Phelan, P.E. et al. (2015) *Urban Heat Island: Mechanisms, Implications, and Possible Remedies*. Annual Review of Environment and Resources 40, H. Volume 40, 2015, S. 285–307. <https://doi.org/10.1146/annurev-environ-102014-021155>
- [3] Deutscher Wetterdienst (2025) *Klimastatusbericht Deutschland Jahr 2024*. Offenbach: DWD, Geschäftsbereich Klima und Umwelt.
- [4] Intergovernmental Panel on Climate Change (IPCC) [Hrsg.] (2023) *Cities, Settlements and Key Infrastructure Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, S. 907–1040.
- [5] Gesetz zur Einsparung von Energie und zur Nutzung erneuerbarer Energien zur Wärme- und Kälteerzeugung in Gebäuden (2024) (Gebäudeenergiegesetz - GEG).
- [6] DIN 4108-2 (2013) *Wärmeschutz und Energie-Einsparung in Gebäuden - Teil_2: Mindestanforderungen an den Wärmeschutz*. Berlin: Beuth Verlag GmbH.
- [7] Deutscher Wetterdienst (2017) *Ortsgenaue Testreferenzjahre von Deutschland für mittlere und extreme Witterungsverhältnisse*. Offenbach: DWD, Geschäftsbereich Klima und Umwelt.
- [8] Deutscher Wetterdienst Wetter und Klima - Deutscher Wetterdienst - Leistungen - Testreferenzjahre (TRY) [online]. <https://www.dwd.de/DE/leistungen/testreferenzjahre/testreferenzjahre.html> [zugegriffen: 21 Okt. 2025]
- [9] Deutscher Wetterdienst (2011) *Aktualisierte und erweiterte Testreferenzjahre von Deutschland für mittlere, extreme und zukünftige Witterungsverhältnisse*. Offenbach: DWD, Geschäftsbereich Klima und Umwelt.
- [10] GEG-Infoportal - TRY 2011. <https://www.bbsr-geg.bund.de/GEGPortal/DE/Praxishilfen/Testreferenzjahre/TRY2011/TRY2011-node.html> [zugegriffen: 21 Okt. 2025]
- [11] Hartner, M. et al. (2024) *Sommerlicher Wärmeschutz der Zukunft*. Proceedings Bauphysiktag in Weimar 2024: Bauphysik in Forschung und Praxis (accepted for publication). <https://doi.org/10.25643/dbt.59920>
- [12] FB Bauphysik / energetische Gebäudeoptimierung Verfahren zur Bemessung des sommerlichen Wärmeschutzes unter Berücksichtigung der zukünftigen Klimaentwicklung - Fachbereich Bauingenieurwesen an der RPTU [online]. <https://bauing.rptu.de/ags/bauphysik/forschung/aktuelle-projekte/verfahren-zur-bemessung-des-sommerlichen->

- waermeschutzes-unter-beruecksichtigung-der-zukuenftigen-klimaentwicklung [zugegriffen: 23 Okt. 2025]
- [13] Tin Kam Ho (1995) *Random decision forests*. Proceedings of 3rd International Conference on Document Analysis and Recognition Montreal, Que., Canada: IEEE Comput. Soc. Press, S. 278–282.
- [14] Pedregosa, F. et al. (2011) *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research 12, H. 85, S. 2825–2830.
- [15] scikit-learn 1.3.2 (2023) RandomForestRegressor. https://github.com/scikit-learn/scikit-learn/blob/1.3.X/sklearn/ensemble/_forest.py.
- [16] Klein; S.A; et. al (2017) TRNSYS 18: A Transient System Simulation Program. Solar Energy Laboratory, University of Wisconsin, Madison, USA.
- [17] Herz, J. et al. (2025) *Using a Machine Learning Algorithm to predict Summer Overheating in Buildings*. Central Europe towards Sustainable Building 2025 IOP Science (zur Veröffentlichung angenommen).
- [18] Belhaouari, S.B. et al. (2024) Knnor-Reg: A Python Package for Oversampling in Imbalanced Regression. Rochester, NY: Social Science Research Network.

Autor:innen

Dipl.-Biophysiker Jakob Herz

herz_jakob@t-online.de

RPTU Kaiserslautern-Landau

Fachgebiet Bauphysik / Energetische Gebäudeoptimierung

Paul-Ehrlich-Straße 29

67663 Kaiserslautern

apl. Prof. Dr. rer. nat. Svenja Carrigan (Korrespondenzautorin)

svenja.carrigan@rptu.de

RPTU Kaiserslautern-Landau

Fachgebiet Bauphysik / Energetische Gebäudeoptimierung

Paul-Ehrlich-Straße 29

67663 Kaiserslautern

Dipl.-Ing. Dipl. Wirt.-Ing Christoph Beecken

c.beecken@bow-ingenieure.de

bow ingenieure gmbh

Hagenmarkt 8

38100 Braunschweig

Prof. Dr. rer. nat. Oliver Kornadt

oliver.kornadt@rptu.de

RPTU Kaiserslautern-Landau

Fachgebiet Bauphysik / Energetische Gebäudeoptimierung

Paul-Ehrlich-Straße 29

67663 Kaiserslautern