

## 2周波GNSS受信機と気象センサを用いた機械学習による降雨予測の検討

Machine learning approach to rainfall nowcast using dual-frequency GNSS device and meteorological sensor

中川 豊<sup>†</sup>                      宮内 大樹<sup>†</sup>                      東野 武史<sup>†</sup>                      岡田 実<sup>†</sup>  
Yutaka Nakagawa              Taiki Miyauchi                      Takeshi Higashino              Minoru Okada

奈良先端科学技術大学院大学 先端科学技術研究科 情報科学領域<sup>†</sup>

### 1 Abstract

The zenith total delay (ZTD) measurement in the GNSS precise positioning is a meteorologically important factor. However, ZTD does not contain enough information to predict future rain when the ZTD is itself employed as a feature quantity. We combined multiple parameters with the ZTD and obtained them from meteorological sensor as explanatory variables, and performing machine learning regression into rainfall intensity. We set the amount of rainfall as the objective variable. Since the GNSS device is for consumer use and it has widely spread into society, it can be expected to be used for disaster prevention network with low cost.

### 2 まえがき

地球大気や海洋・陸地の状態の変化を計算機シミュレーションによって予測する数値予報 [1] による降雨量の予報では、大気を連続量として扱い、離散化した格子点を観測点とし、大気の変化を表す方程式によりその過程を算出することで予測を行っている。観測値の初期値は主に気象レーダや気象衛星などのリモートセンシング技術で得られる。気象レーダは大型で広範囲（数百 km レンジ）をカバーするように設計されているが、地球のまるみに隠れた場所ではレーダのビームが地表から離れることで、地表付近を観測できない。ところが、積乱雲が引き起こす強雨や竜巻などの現象は地表付近で起こるため、地表付近の観測を密にすることが求められている。気象衛星は水平距離の分解能が 2km と向上しているが、豪雨の予兆の把握に必要となる大気の中・下層の水蒸気の観測精度は十分ではない。そこで GNSS(global navigation satellite system) ゾンデがこれらを補完する候補として挙げられる。測位衛星からの電波は地上に届くまでの対流圏内で水蒸気の影響を受け余分な遅延時間を生じる。この現象を雨雲が発生する前兆として捉え、降雨の予測に応用できる。GNSS 受信機と測位解析ソフトウェアによって、GNSS 測位情報から対流圏遅延量がリアルタイムに推定できる。様々な精密測位方式のなかでも、基準局を必要としない精密単独測位法 (PPP: precise point positioning) は、測位精度が相対測位精度程度に高精度であれば、最も望ましい測位方法といわれている。PPP を用いて観測点上空における水蒸気による伝搬遅延量から天頂遅延量 (ZTD: zenith total delay) を抽出することで水蒸気量を見積もることができる。しかし水蒸気

の指標となる ZTD 単独で降雨の有無判定を行っても警報誤りが多いことがわかった [2]。また、地上の気象センサから得られる降雨強度、気温、湿度、気圧等は降雨強度に対する相関が非常に小さいことがわかった。そこで我々は機械学習のモデルの 1 つであるランダムフォレストにより降雨強度を回帰してその精度の評価を行った [3]。機械学習にはランダムフォレスト以外にも様々な方法があり、特性に応じた最適なモデルを選択する必要がある。本稿では、データの特性に応じた最適なモデル選択を行うことを目的とし、夏季 (6 月) と冬季 (2 月) の 2 つの時期の観測データを様々な機械学習モデルを用いて評価した結果を報告する。観測データの蓄積とモデルの精度の向上により大雨の予測精度向上が期待できる。

### 3 観測システム

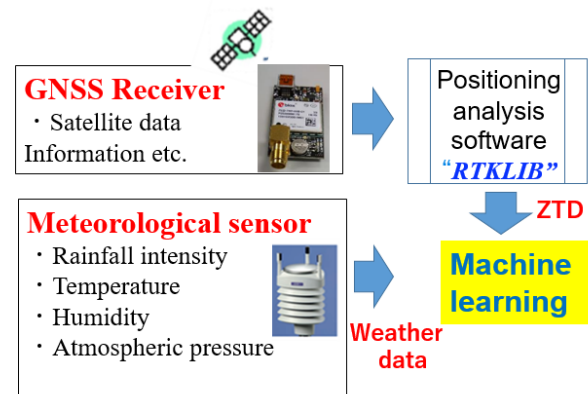


図 1 システム構成

図 1 にシステム構成を示す。GNSS のアンテナとして Ashtech 製 ASH701941.B(チョークリングコイル型)を用いた。GNSS 受信器として u-blox 製 F9P を用いて衛星信号を受信し、デコードの後ログに記録する。対応する衛星は GPS/GLO であり、使用する GNSS 信号は、GPS:L1C/A 及び L2C と GLONASS:L1OF 及び L2OF とした。本実験で用いる GNSS 受信機は 2 周波に対応しており、測位誤差の要因として最も大きなウエイトを占める電離層の遅延による影響を軽減することで、精度のよい ZTD が取得できる。測位解析用のソフトウェアとして、オープンソースの RTKLIB[4] を用いた。これによりタイムスタンプ付きの ZTD を取得するためのポジショニングプロセスを監視できる。

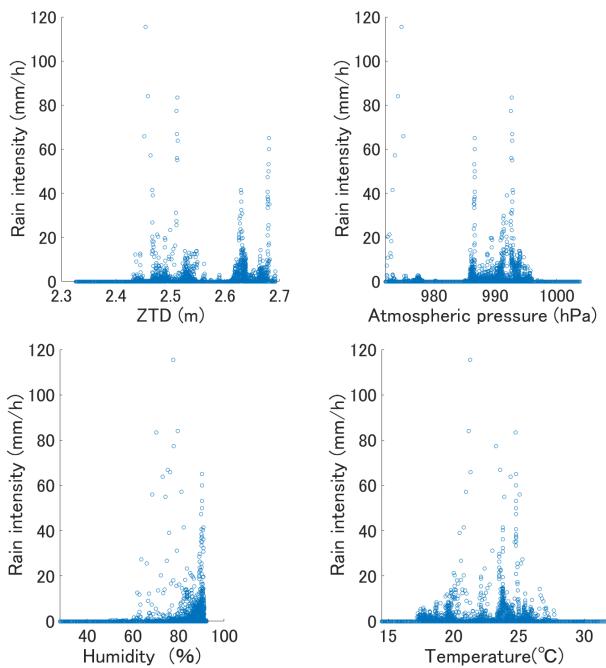


図 2 ZTD, 気象センサの各要素と降雨強度の関係 (Jun 2019)

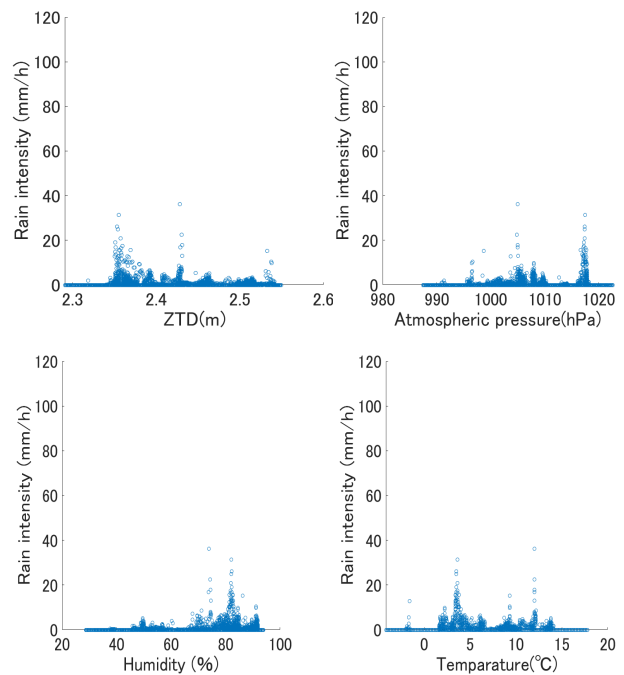


図 3 ZTD, 気象センサの各要素と降雨強度の関係 (Feb 2020)

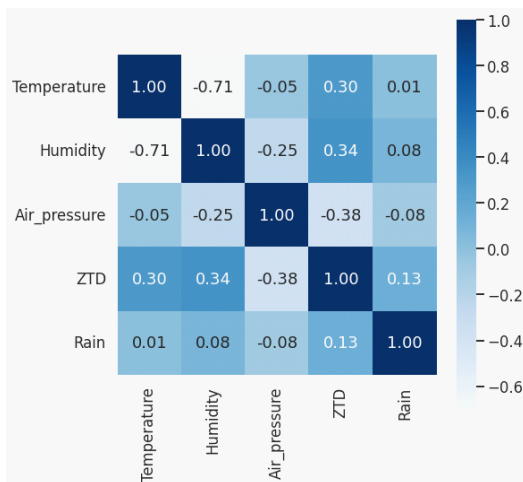


図 4 ZTD, 気象センサの各要素の相関 (Jun 2019)

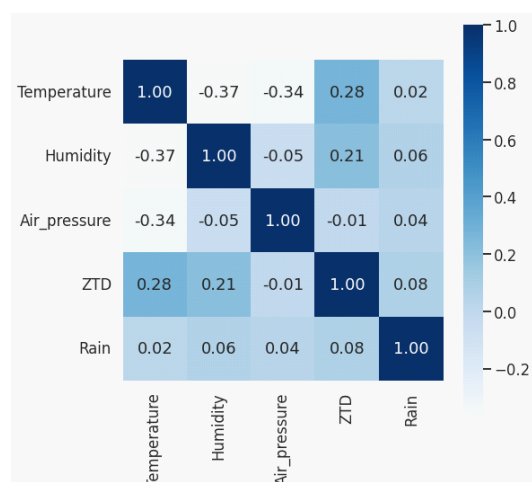


図 5 ZTD, 気象センサの各要素の相関 (Feb 2020)

降雨強度や気温などは奈良先端科学技術大学院大学<sup>1</sup>構内の気象センサを利用した。取得した ZTD と気象センサの時系列ログはデータフレームに結合され、機械学習ベースの回帰のパラメータに使用される。図 2、図 3 にそれぞれ 2019 年 6 月および 2020 年 2 月の降雨強度に対する ZTD と気象センサの各要素の散布図を示す。図に示すように ZTD と各種気象要素に比べて降雨強度は大きく振れる。また、降雨強度が 0 となるサンプルが大部分を占めている。なお、観測期間を通じて相対湿度は 94% 以下であった。降雨時の相対湿度の観測値が 100% にならないことは、地表付近と上空の湿度が異なること及び観測値に含まれる  $\pm 5\%$  の絶対誤差に起因する。

<sup>1</sup>奈良県生駒市

図 4、図 5 はそれぞれ 2019 年 6 月と 2020 年 2 月の ZTD および気象センサの各要素間の相関を示す。図 4、図 5 の *Rain* は降雨強度 [mm/h] を示しており、降雨強度に対する各要素との相関係数は 2019 年 6 月の場合  $-0.08$  から  $0.13$  で、2020 年 2 月では  $0.02$  から  $0.08$  に分布している。相関係数の絶対値の平均は 7 月と 2 月でそれぞれ  $0.233$ 、 $0.146$  であり、7 月の方が要素間の相関が高いことがわかる。また、両月とも ZTD との相関が最も高いことがわかった。また図 2、図 3 および図 4、図 5 から各気象要素と降雨強度の間の相関もほとんどないことがわかる。つまり降雨強度と関係性がある要素を簡単には見い出せないため ZTD と気象センサを組み合わせさせたデータを機械学習へ適用する動機となっている。

## 4 機械学習のモデルの比較

機械学習では、既知の入力および出力データをモデルに学習させて将来の出力を予測できるようにする教師あり学習と、入力データの隠されたパターンまたは内在する構造を探す教師なし学習という 2 種類の手法があるが、本稿では教師ありの機械学習を対象としている。その目的は、不確定要素が存在する状況に対して、正解が明確である証拠に基づいて予測を行うモデルを構築することである。教師あり学習アルゴリズムでは、既知の入力データのセットとそのデータに対する既知の応答 (出力) を使用して、新しいデータに対する応答を適切に予測するようモデルに学習をさせる。教師あり学習では、回帰などの手法を使用して予測モデルを構築する。機械学習アルゴリズムには多くの種類があり、それぞれが異なるアプローチで学習を行うので、最適な方法や汎用の方法は簡単に判断できない。適切なアルゴリズムを見つけることは、ある程度は試行錯誤に基づく。正しいアルゴリズムを選択するには、モデルの速度、精度、複雑さなど、あるメリットと別のメリットのトレードオフが必要になる。本研究ではいくつかのアルゴリズムで機械学習を行いその精度を比較検証して最適な予測モデルを選択する。

### 4.1 機械学習モデル

本研究では降雨強度の回帰に以下のモデルを用いた。

1. Linear regression model  
一般化線形モデル (GLM)
2. Linear interaction model  
GLM と説明変数モデルに交互作用項を含めたモデル
3. Bagging tree model  
ブートストラップによるアンサンブル学習モデル
4. Boosting tree model  
ブースティングによるアンサンブル学習モデル
5. Coarse tree model  
大きな葉の応答関数の柔軟性が低い決定木モデル
6. Medium tree model  
普通のサイズの葉の応答関数の柔軟性が低い決定木モデル
7. Complex tree model  
多くの小さな葉で応答機能の柔軟性が高い決定木モデル
8. Fine Gaussian SVM model  
応答関数が柔軟なサポートベクターマシン (SVM)
9. Narrow neural network model  
ニューラルネットワークモデル  
(レイヤー数：1, レイヤーサイズ：10, ReLU 関数)
10. Medium neural network model  
ニューラルネットワークモデル  
(レイヤー数：1, レイヤーサイズ：25, ReLU 関数)

11. Wide neural network model  
ニューラルネットワークモデル  
(レイヤー数：1, レイヤーサイズ：100, ReLU 関数)
12. 2-layer neural network model  
2 層ニューラルネットワークモデル  
(レイヤー数：2, レイヤーサイズ：10, ReLU 関数)
13. 3-layer neural network model  
3 層ニューラルネットワークモデル  
(レイヤー数：3, レイヤーサイズ：10, ReLU 関数)
14. Squared index GPR model  
二乗指数のガウス過程回帰 (GPR) モデル

### 4.2 テスト方法

機械学習のテスト方法は一般にホールドアウト法と交差検証法がある。ホールドアウト法とはデータを訓練 (トレーニング) データと検証用のテストデータに分割して評価を行う方法である。ホールドアウト法はデータの分割の仕方により、評価値が分散する可能性があることが問題とされている。一方で交差検証法 (Cross validation) では、データを  $m$  個のグループに分割し、 $m - 1$  個のグループのデータを使って学習し、残りの 1 つのグループを検証する。ホールドアウト法の問題点となるばらつきの大きなデータの場合でも使用可能な方法であるため、本稿でも交差検証法を用いた。以下に手順を示す。

1. 全データを 3 分割以上の任意の数 (本稿では  $m=5$ ) に分割する。
2. 上の分割から 4 つのグループをトレーニングデータとして選び、回帰モデルを生成する。
3. 残りの 1 つをテストデータとして回帰モデルに入力する。
4. 出力結果とテストデータの正答を比較する。
5. 以上をプロセスを分割数だけ繰り返す。
6. 全てのプロセスが完了したら、評価値の平均を算出する。

## 5 評価

### 5.1 評価基準

回帰モデルを評価する際は、実測値と予測値の差 (残差) を使った指標を使うのが一般的である。各モデルによる評価を表 1 表 2 および図 6, 図 7 に示す。評価のための指標は RSME と  $R^2$  を用いた。RMSE とは、Root Mean Square Error (二乗平均平方根誤差) の略で、回帰モデルの最も一般的な性能指標で以下の式 (1) で表現される。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

ここで、 $y_i$ : 真値  $\hat{y}_i$ : 予測値である。

RMSE は「真値と予測値の誤差の二乗の総和の平均値の平方根」であり、RMSE が小さいほど機械学習モデルの性能が良いと判断できる RMSE は一般的な性能評価指標であり、多くの場合は RMSE が使われる。概ねの回帰タスクに適しており大きな誤差に着目したい場合には RMSE が適している。なお、真値と予測値の誤差を二乗するため、頻度は少ないが大きな誤差があった場合に影響を受けやすい。一方で決定係数  $R^2$  は、以下の式 (2) で表される。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2)$$

ここで  $\bar{y}_i$  は真値の平均を表す。 $R^2$  は回帰によって導いたモデルの当てはまりの良さを表現する値で、予測値が真値とどの程度一致するかを表す。決定係数  $R^2$  は 0 から 1 の範囲の値をとり、値が大きいほどモデルが適切にデータを表現している。一般に決定係数  $R^2$  の判断目安は文献 [5] などに示す、以下のような例が多い。

1. 0.5 以下：  
モデルとして意味をなさない。
2. 0.8 以上：  
良いモデルと考えられる。ただ 0.8 以下の数値であってもデータ同士の関連性が存在する場合もあるので一つの指標としてみる必要がある。
3. 0.9 以上：  
過学習の有無に対する注意が必要である。

ここで式 (1),(2) から RMSE と  $R^2$  の関係を明らかにしておく。 $|y_i - \hat{y}_i|$  が小さい場合、予測誤差が小さいことを表す。すなわち RMSE が小さく、 $R^2$  が大きい (1 に近づく) ことを表す。これに対し  $|y_i - \hat{y}_i|$  が大きい場合は予測誤差が大きくなることで、RMSE が大きくなり  $R^2$  が小さい (0 に近づく) ことを表す。また、極端に誤差が大きい評価値が含まれると RMSE が大きくなりやすい。

## 5.2 評価の結果

今回の評価では計 14 のモデルの比較を行った。表 1, 表 2 および図 6, 図 7 に各モデルの評価の結果を示す。表 3 には 2 つの事例について RMSE と  $R^2$  について基本統計量を示した。RMSE については 2019 年 6 月の方が 2020 年 2 月に比べて全体的に大きくなった。6 月は梅雨の時期であり、大雨が発生しやすく降雨強度の変化量が激しくなる。降雨強度のバラツキが大きくなりその結果、RMSE が大きくなったことも要因と推測できる。個々のモデルの検証では、両月共に Bagging tree, Complex tree, Squared index GPR の 3 つが RMSE が低くかつ  $R^2$  が高く、モデルへの適正として良い結果となった。3 節で述べたように 6 月は要素間の相関が 2 月よりも高い特徴がある。相関の高い変数を教師データとする場合、特に決定木を用いたアンサンブル学習では、決定木の構造の相関が高くなることで決定木間の出力が似かより十分な性能強化が得られないことがある。しか

し、今回の評価では Bagging Tree の評価値が Boosting tree よりも高いため、決定木間の相関はアンサンブル学習の性能を劣化させるほど高くないことがわかる。さらに、6 月から 2 月の評価値に改善があることから本システムで取得した ZTD および気象要素変数の相関を低減すればより高い汎化能力の回帰が可能であることがわかる。

表 1 個々のモデルの評価 (Jun 2019)

Model	RMSE	$R^2$
Linear	1.7395	0.02
Linear interaction	1.7282	0.03
Bagging tree	0.92517	0.72
Boosting tree	1.3458	0.41
Coarse tree	1.4091	0.36
Medium tree	1.1788	0.55
Complex tree	0.79647	0.79
Fine Gaussian SVM	1.627	0.14
Narrow neural network	1.6032	0.17
Medium neural network	1.5684	0.2
Wide neural network	1.3941	0.37
2-layer neural network	1.254	0.49
3-layer neural network	1.2091	0.53
Squared index GPR	0.71194	0.84

表 2 個々のモデルの評価 (Feb 2020)

Model	RMSE	$R^2$
Linear	0.67457	0.01
Linear interaction	0.67162	0.02
Bagging tree	0.27704	0.83
Boosting tree	0.43728	0.58
Coarse tree	0.43065	0.6
Medium tree	0.32331	0.77
Complex tree	0.22732	0.89
Fine Gaussian SVM	0.50884	0.44
Narrow neural network	0.59885	0.22
Medium neural network	0.55545	0.33
Wide neural network	0.45687	0.55
2-layer neural network	0.44156	0.58
3-layer neural network	0.34565	0.74
Squared index GPR	0.31338	0.79

GPR(Squared index GPR) による回帰が良好であった理由は、各要素の確率分布ならびに降水強度と気象要素の同時分布、そして条件付き確率がガウス分布によく近似されたためだと考えられる。GPR においても、ZTD と気象要素間の相関が低い 2 月のデータの方が評価値は良好であり、気象観測変数間の無相関化の操作が必要になることが明らかとなった。個々の気象要素の分布のフィッティングについては今後の課題としたい。

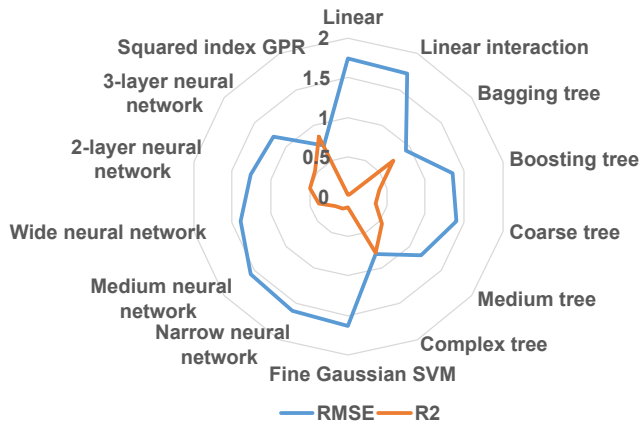


図 6 個々のモデルの評価 (Jun 2019)

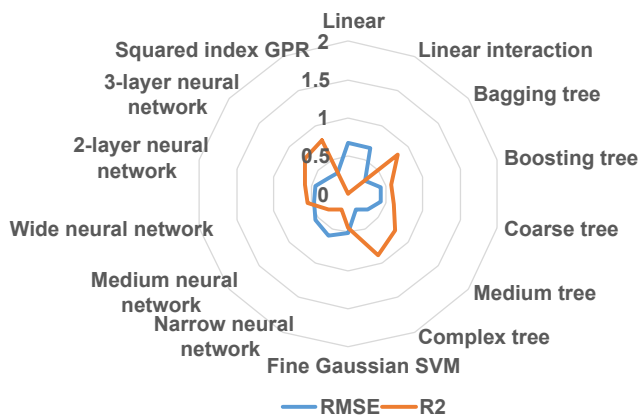


図 7 モデルの評価 (Feb 2020)

表 3 Jun 2019 と Feb 2020 の統計量の比較

	Jun 2019		Feb 2020	
	RMSE	$R^2$	RMSE	$R^2$
平均	1.32077	0.401429	0.447314	0.525
標準誤差	0.088614	0.071866	0.037938	0.076803
中央値	1.36995	0.39	0.43942	0.58
標準偏差	0.331564	0.268897	0.141949	0.287369
分散	0.109935	0.072305	0.02015	0.082581
最小	0.71194	0.02	0.22732	0.01
最大	1.7395	0.84	0.67457	0.89

図 8、図 9 に 2019 年 6 月の Linear regression model と 2020 年 2 月の Complex tree model の予測結果の例を示す。それぞれ縦軸は予測値、横軸は真値である。図 8 は全く予測ができていない例である。対して図 9 は降雨強度が大きくなるほどバラツキが大きくなるが全モデルの中で最も良いスコアの例である。

## 6 まとめ

一般に予測モデルは回帰問題を基本とするが、説明変数 (ZTD と降雨強度を除いた気象要素) と目的変数 (降雨強度) にほとんど相関がみられないため、機械学習モデルによる降雨強度の回帰を行った。

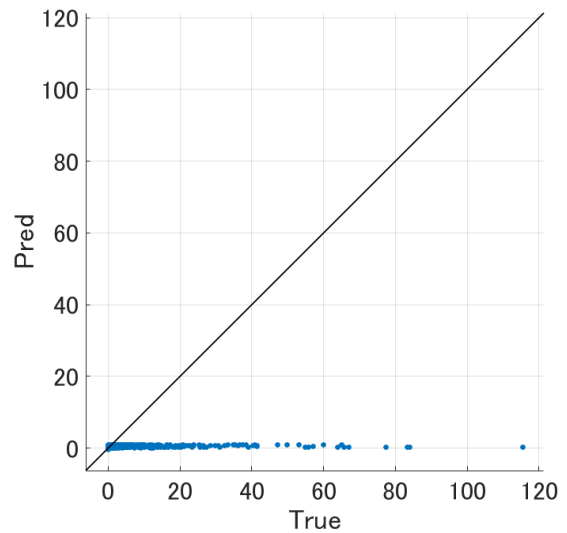


図 8 Linear regression model(2019 年 6 月)

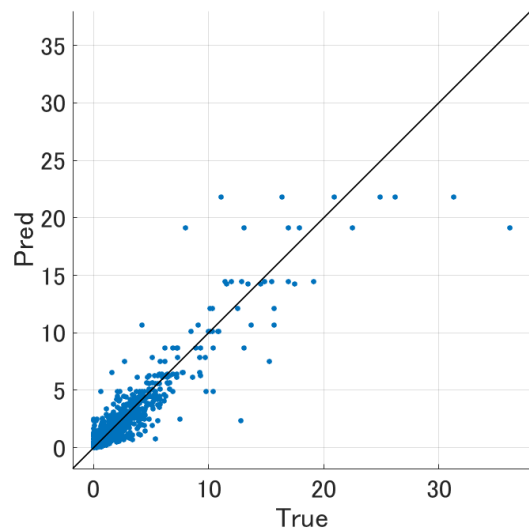


図 9 Complex tree model(2020 年 2 月)

機械学習では、既知の入力および出力データをモデルに学習させて将来の出力を予測できるようにする教師あり学習と、入力データの隠されたパターンまたは内在する構造を探す教師なし学習という 2 種類の手法があるが、本稿では不確定要素が存在する状況に対して、正解が明確である証拠に基づいて予測を行うモデルを構築することを目的として教師ありの機械学習を対象とした。教師ありの機械学習では過学習に注意が必要であるが、K 分割交差検証を用いることでその対策とした。本研究において、様々な機械学習モデルの中から降雨強度の予測性能の高いモデルを確認できた。今後の課題は過学習の検証と特徴量の種々選択である。ZTD や気象要素のほか新たに時間変化量や相互作用などの特徴量を追加することとデータの蓄積により、予測精度が向上できれば大雨に対する防災への貢献が期待できると考える。

## 参考文献

- [1] 木田秀次, “我国の数値予報システムの現状 (1) 概要,” 天気, vol.37, no.9, pp.p591–602, 1990.
- [2] 中川豊ほか, “衛星測位情報を応用した降雨予測の検討,” 電気関係学会関西連合大会論文, pp.166–167, 2019.
- [3] Y. Nakagawa, et al., “An application of random forest regression for rainfall intensity estimation using gnss receiver,” 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE 2021), pp.148–149, 2021.
- [4] T. Takasu, N. Kubo, and A. Yasuda, “Development, evaluation and application of rtklib: a program library for rtk-gps,” GPS/GNSS symposium, pp.213–218, 2007.
- [5] 児玉紘幸, “データマイニングを活用したモノづくりの意思決定支援,” 精密工学会誌, vol.83, no.11, pp.1014–1017, 2017.