

PERBANDINGAN MODEL KALIBRASI BERBASIS *PLASMA-ACTIVATED WATER* MENGGUNAKAN *PRINCIPAL COMPONENT REGRESSION* DAN *PARTIAL LEAST SQUARE REGRESSION* DALAM R

Suhartono ^{1*}, Muhammad Subianto ²

¹Jurusan Teknik Informatika, FST, UIN Maulana Malik Ibrahim Malang, Jawa Timur, Indonesia.

²Jurusan Informatika, FMIPA, Universitas Syiah Kuala, Aceh, Indonesia.

e-mail: ^{1*} suhartono@ti.uin-malang.ac.id

*Penulis Korespondensi

Diserahkan: 18-10-2024; Direvisi: 25-10-2024; Diterima: 22-11-2024

Abstrak: Untuk memastikan alat reaktor *plasma* dapat mensimulasikan larutan *Plasma Activated Water* (PAW) secara akurat. Untuk mendapatkan keterjaminan kualitas dari larutan PAW sesuai dengan desain reaktor *plasma*, maka perlu dibuat model kalibrasi yang dapat memastikan bahwa model sesuai dengan data pengamatan, dan meningkatkan kemampuan prediktif model. Tujuan artikel ini adalah untuk membangun model kalibrasi berdasarkan kualitas larutan PAW yang dihasilkan dari reaktor *plasma*. Pada penelitian ini menggunakan metode *Principal Component Regression* (PCR) dan metode *Partial Least Square Regression* (PLS-R). Kelebihan metode PCR adalah melakukan reduksi data berdasarkan nilai korelasi. Sedangkan metode PLS-R melakukan reduksi data berdasarkan faktor-faktor yang paling relevan dalam menginterpretasi data. Berdasarkan percobaan yang dilakukan disimpulkan bahwa untuk membangun model kalibrasi berdasarkan data reaktor *plasma* dan data larutan PAW maka metode PCR lebih baik dibanding metode PLS-R. Hal ini ditunjukkan berdasarkan nilai RMSEP dan R^2 pada metode PCR sebesar 0,09625571 dan 93,04699% sedangkan pada metode PLS-R sebesar 0,09873341 dan 92,84436%. Untuk nilai R^2 pada metode PCR lebih besar yang menandakan bahwa nilai varian data lebih dapat diterima oleh model kalibrasi dibanding pada metode PLS-R, kemudian nilai RMSEP pada metode PCR lebih kecil yang menandakan nilai *error* statistik lebih dapat diterima dibanding dengan PLS-R.

Kata Kunci: model kalibrasi; *principal component regression*; *partial least squares regression*; prediksi; reaktor *plasma*, larutan

Abstract: To ensure that the plasma reactor tool can simulate the Plasma Activated Water (PAW) liquid accurately. To ensure the quality of the PAW liquid according to the plasma reactor design, it is necessary to create a calibration model that can ensure that the model matches the observation data and improve the predictive ability of the model. The purpose of this article is to build a calibration model based on the quality of PAW liquid produced from a plasma reactor. This study used the Principal Component Regression (PCR) method and the Partial Least Square Regression (PLS-R) method. The advantage of the PCR method is that it reduces data based on correlation values. While the PLS-R method reduces data based on the most relevant factors in interpreting the data. Based on the experiments conducted, it was concluded that to build a calibration model based on plasma reactor data and PAW liquid data, the PCR method is better than the PLS-R method. This is shown based on the RMSEP and R^2 values in the PCR method of 0.09625571 and 93.04699% while in the PLS-R method of 0.09873341 and 92.84436%. For the R^2 value in the PCR method is greater which indicates that the data variant value is more acceptable to the calibration model than in the PLS-R method, then the RMSEP value in the PCR method is smaller which indicates that the statistical error value is more acceptable than PLS-R.

Keywords: calibration model; principal component regression; partial least squares regression; prediction; plasma reactor; liquid

Kutipan: Suhartono & Subianto, Muhammad. (2025). Perbandingan Model Kalibrasi Berbasis *Plasma-Activated Water* Menggunakan *Principal Component Regression* dan *Partial Least Square Regression* Dalam R. *JP2M (Jurnal Pendidikan dan Pembelajaran Matematika)*, Vol.11 No.1, (11-26). <https://doi.org/10.29100/jp2m.v11i1.6869>



Pendahuluan

Model kalibrasi sangat penting dalam berbagai bidang penelitian dan aplikasi, yang berfungsi untuk meningkatkan akurasi dan keandalan model prediktif. Kebutuhan untuk membangun model kalibrasi karena ketidakpastian dan variabilitas inheren yang ada dalam data pengamatan dan model kalibrasi mencoba untuk mewakili dalam semua kondisi. Model kalibrasi adalah fungsi yang menggambarkan hubungan antara variabel-variabel pada data pengamatan, dimana proses pada model kalibrasi menyeimbangkan berbagai metrik kinerja, dan memastikan bahwa model adalah tangguh di berbagai scenario (Tuo et al., 2018; Sung & Tuo, 2024).

Sedangkan pada data pengamatan terdapat beberapa masalah yang harus diselesaikan sebelum kita melakukan proses model kalibrasi, ketersediaan data pengamatan adalah jarang, dimensi data tidak konsisten, atau data dikumpulkan pada interval yang tidak teratur, yang dapat menyebabkan estimasi parameter yang bias dan mengakibatkan kinerja model yang buruk (Mai, 2023).

Pada beberapa penelitian telah menunjukkan bahwa data tingkat aliran yang bersumber dari banyak orang sering kali mengalami interval waktu yang tidak teratur dan resolusi vertikal yang terbatas, sehingga mempersulit proses kalibrasi (Van Meerveld et al., 2017). Pada penelitian kalibrasi model bidang hidrologi, data pengamatan yang tersedia tidak secara memadai mewakili variabilitas sistem yang dimodelkan, yang dapat menyebabkan overfitting, di mana model kalibrasi berkinerja baik pada set data training tetapi buruk pada set data testing (Da Silva et al., 2018).

Untuk menyelesaikan permasalahan pada data pengamatan, salah satunya adalah terdiri dari banyak dimensi, pertama, kita harus melakukan analisis untuk mengurangi dimensi guna meningkatkan kinerja model kalibrasi. Yang kedua adalah menggunakan teknik reduksi data yang digunakan untuk membangun model kalibrasi, reduksi data merupakan teknik penting dalam kalibrasi model, khususnya saat menangani sistem kompleks yang dicirikan oleh kumpulan data berdimensi tinggi (Waldron et al., 2022). Proses ini melibatkan penyederhanaan data dengan mengurangi jumlah variabel atau komponen sambil mempertahankan informasi penting yang diperlukan untuk kalibrasi model yang efektif. Keuntungan signifikan lainnya dari reduksi data adalah mitigasi overfitting. Kumpulan data berdimensi tinggi sering kali menghasilkan model yang sangat cocok dengan data pelatihan tetapi berkinerja buruk pada data yang tidak terlihat karena kompleksitasnya.

Dengan mengurangi jumlah komponen, peneliti dapat menyederhanakan struktur model, sehingga meningkatkan generalisasinya (Wijewardane et al., 2021). Pada penelitian terkait dengan komunikasi hasil yang lebih baik kepada audiens non-teknis dengan cara mengurangi jumlah parameter dalam model kalibrasi untuk membantu memperjelas faktor mana yang paling berpengaruh dalam mendorong hasil yang diinginkan (Saleh et al., 2020). Untuk membangun model kalibrasi yang terbaik, pada penelitian ini menggunakan dua metode dalam menangani proses reduksi, oleh karena itu pada artikel ini menjelaskan komparasi model kalibrasi pada reaktor *plasma* menggunakan *principal component regression* (PCR) dan *partial least square regression* (PLS-R) dalam bahasa pemrograman R.

Teknik seperti Analisis Komponen Utama (PCA) atau metode pemilihan fitur dapat digunakan untuk mengidentifikasi dan mempertahankan hanya variabel yang paling informatif, yang membantu dalam membangun model yang lebih kuat (Pianosi & Wagener, 2016). Metode PCR dapat menjelaskan variabilitas data pengamatan dengan kombinasi linier yang ditemukan dari berbagai variabel independen

yang dapat digunakan untuk mengurangi komponen utama data kalibrasi (Ahmed & Kim, 2018). Metode PCR adalah sangat efektif ketika tujuannya adalah untuk memvisualisasikan data atau mengurangi gangguan sambil mempertahankan varians paling signifikan dalam himpunan data (Malik, 2015; He et al., 2005). Namun, PCA tidak mempertimbangkan variabel respons, yang dapat membatasi efektivitasnya dalam skenario pemodelan prediktif di mana hubungan antara prediktor dan hasil sangat penting (M. et al., 2020).

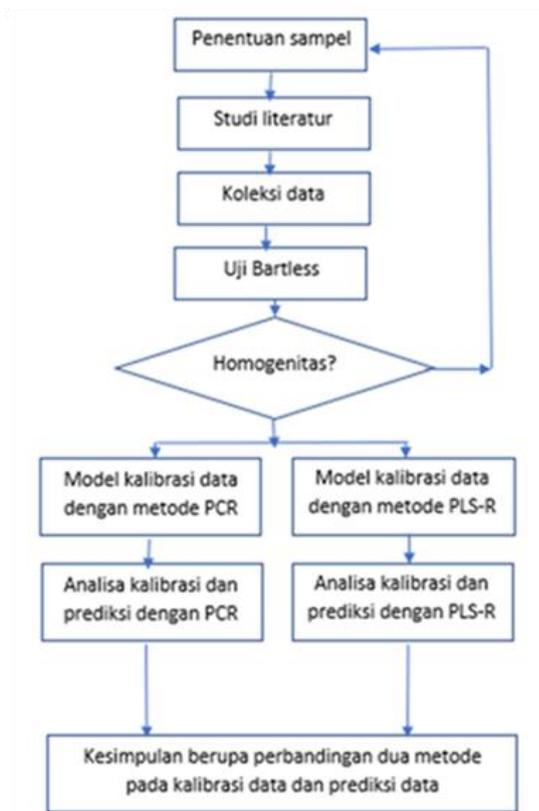
Dalam praktiknya, berbagai metode dapat diterapkan untuk reduksi komponen data dalam model kalibrasi. Misalnya, pendekatan metode *Partial Least Square Regression* (PLS-R) dapat digunakan untuk menggabungkan pengukuran eksperimental dengan prediksi model, yang secara efektif mengurangi dimensionalitas data sekaligus menjaga integritas proses kalibrasi (Rasmussen et al., 2017). Metode PLS-R digunakan sebagai pendekatan reduksi data sekaligus sebagai cara untuk menentukan aspek yang paling relevan dalam peramalan dan interpretasi data. PLS-R berkonsentrasi pada kovarians respon dan variabel predikto (Mevik & Wehrens, 2007). PLS-R dirancang untuk menangani situasi di mana jumlah prediktor melebihi jumlah observasi, membuatnya sangat berguna dalam bidang seperti kemometrika dan spektroskopi (Rácz et al., 2015 ; Singh et al., 2013). PLS-R tidak hanya mengurangi dimensionalitas tetapi juga memodelkan hubungan antara prediktor dan variabel respons secara langsung, yang meningkatkan kemampuan prediktifnya (Tan et al., 2008).

Dengan membandingkan penggunaan metode PCR dan metode PLS-R untuk membangun model kalibrasi dari reaktor plasma maka kita dapat menyeimbangkan informasi antara prediktor dan respons, serta efektivitas kedua metode, prediksi dapat ditingkatkan, terutama untuk menghasilkan larutan PAW yang baik berdasarkan berbagai data dari variabel pada sensor dan kontrol dalam reaktor *plasma*. Larutan PAW adalah larutan air yang telah diolah dengan aliran gas terionisasi atau disebut *plasma* (Zhao et al., 2020 ; Thirumdas et al., 2018).

Metode

Tahapan penelitian

Pada gambar 1 menjelaskan tahapan penelitian dalam bentuk diagram penelitian. Tahap pertama data pengamatan dilakukan uji homogenitas, tahapan kedua setelah lolos uji homogenitas adalah membangun model kalibrasi dengan menggunakan metode PCR dan PLS-R, model kalibrasi digunakan untuk memprediksi larutan PAW yang dihasilkan dari reaktor *plasma*. Tahapan ketiga adalah membandingkan kedua metode pada model kalibrasi yang telah dibangun, tahapan keempat adalah menentukan metode yang terbaik berdasarkan data pengamatan.



Gambar 1. Diagram tahapan penelitian

Perangkat lunak yang digunakan

Perangkat lunak *open source* pemrograman R digunakan dalam penelitian ini untuk melakukan evaluasi kualitas, pemodelan, dan pengujian signifikansi statistik. Paket pada pemrograman R *open-source* yang digunakan adalah paket `pcr()` dan `pls()`. Kedua fungsi tersebut telah digunakan untuk menentukan jumlah komponen yang akan dipertahankan untuk menghindari *overfitting* karena terlalu banyak prediktor yang disertakan (Buker et al., 2022). Penggunaan dua paket perangkat lunak digunakan untuk melakukan berbagai analisis, membuat grafik, dan menyiapkan perhitungan untuk publikasi. Paket `pcr()` dan `pls()` digunakan dalam membangun model kalibrasi, dan keduanya diterapkan pada data pengamatan yang diperoleh dari reaktor *plasma*. Selain itu, penelitian ini melakukan analisis komparasi untuk menganalisis hasil yang diperoleh dari dua pendekatan metode untuk memastikan keandalannya.

Pengumpulan data

Data pengamatan dikumpulkan dari reaktor *plasma* berdasarkan sensor dan perangkat keras yang terpasang pada Arduino untuk pengukuran. Pengambilan data meliputi data suhu ruangan, tegangan tegangan tinggi, data kecepatan motor, dan data *ultraviolet*. Ke-empat data tersebut digunakan sebagai variabel masukan. Skrip kode R, yang dapat dilihat pada gambar 2, digunakan untuk melihat *dataset*. Pada gambar 3 menunjukkan *output* dari kode skrip R proses pengambilan data dari *file excel* dan menampilkan dalam format tabel.

```

1  "" {r}
2  ## Import the data
3  library(readxl)
4  seer_ordinal <- read_excel(path = "sensorup2.xlsx")
5  seer_ordinal
6  ""
  
```

Gambar 2. Kode skrip R untuk mengambil data file excel `sensorup2.xlsx`

PAW	SPEKTROFOTOMETRI ULTRAVIOLET	Sensor_Cahaya	Suhu	Voltase KV	Kipas
1.70	0.5674	0.14	26.9	600	400
1.15	0.3681	0.03	25.2	200	125
1.27	0.3714	0.05	26.8	200	175
1.29	0.3806	0.06	27.8	200	200
1.10	0.3611	0.02	26.1	200	100
1.34	0.4632	0.08	27.9	400	250
1.39	0.4835	0.09	28.3	400	275
1.50	0.4739	0.10	29.2	400	300
2.30	0.6475	1.24	30.7	800	550
1.65	0.5361	0.12	29.2	600	350

Gambar 3. Hasil kode skrip R untuk mengambil data file excel dan menampilkan dalam bentuk tabel

Bagaimana data pengukuran dari sensor, yang meliputi data suhu ruangan dalam reaktor *plasma*, data tegangan tinggi, data kecepatan motor, sensor cahaya, dan sensor *ultraviolet*, dapat diambil dan ditampilkan dalam bentuk tabel. Sedangkan, variabel keluaran adalah data larutan PAW berdasarkan pengukuran laboratorium. Data pengamatan dibagi menjadi dua *dataset* yaitu *train.data* adalah data yang digunakan untuk melatih model kalibrasi dan *test.data* adalah data yang digunakan untuk menguji akurasi model kalibrasi. Pada gambar 4 menjelaskan pembagian data menjadi *train.data* dan *test.data* menggunakan kode skrip R. Data pengamatan terdiri dari variabel output dari data larutan PAW yang dievaluasi di laboratorium. Sedangkan sebagai variabel input sebanyak lima data dari reaktor *plasma* yang ditangkap oleh arduino, kedua variabel tersebut kemudian digabung menjadi satu *dataset*, *dataset* tersebut digunakan untuk membangun model kalibrasi dengan menggunakan metode PCR dan PLS-R.

```

1  "" {r}
2  train.data <- seer_ordinal[1:10,]
3  test.data <- seer_ordinal[11:20,]
4  ""
    
```

Gambar 4. Kode skrip R untuk membagi data menjadi *train.data* dan *test.data*

Testing data pengamatan

Sebelum menerapkan metode PCR dan PLS-R untuk membangun model kalibrasi, maka data pengamatan harus dilakukan uji Bartlett (Pandey et al., 1983) untuk menentukan apakah *k sampel* data pengamatan memiliki *varians* yang sama. Uji Bartlett adalah uji statistik yang digunakan untuk menentukan apakah varians antara beberapa kelompok sama atau tidak (Arsham & Lovric, 2011). Kode skrip R untuk uji Bartlett digunakan agar data pengamatan memiliki *varians* yang sama di seluruh sampel, yang dikenal sebagai *homogenitas varians*. Gambar 5 menunjukkan pengujian data menggunakan program *EFAtools*. Hasil kode skrip R untuk uji Bartlett *train.data* dengan menggunakan *package EFAtools* dapat dilihat pada gambar 6. Uji Bartlett diklaim signifikan pada tingkat *alfa* 0,05 pada gambar 6, dan data kalibrasi dikatakan sangat sesuai untuk analisis faktor dan dapat digunakan baik untuk metode PCR maupun PLS-R.

```

1  "" {r}
2  library(EFAtools)
3  BARTLETT(train.data , N= nrow(train.data ))
4  ""
    
```

Gambar 5. Kode skrip R untuk uji Bartlett *train.data*

✓ The Bartlett's test of sphericity was significant at an alpha level of .05.
 These data are probably suitable for factor analysis.
 $\chi^2(15) = 98.41, p < .001$

Gambar 6. Hasil kode skrip R untuk uji Bartlett *train.data*

Evaluasi model kalibrasi

Untuk mengetahui kinerja model kalibrasi adalah menggunakan nilai *Root Mean Square Error of Prediction* (RMSEP) dan nilai korelasi R^2 berdasarkan pada data estimasi dan data aktual pada beberapa sampel larutan PAW, hal ini digunakan untuk membandingkan kedua pendekatan tersebut. Penggunaan RMSEP dan R^2 pernah digunakan untuk mengevaluasi kinerja kalibrasi pada model hidrodinamika danau 3D (Baracchini et al., 2020). Nilai RMSEP di dapat dengan menggunakan rumus (1). Sedangkan nilai R^2 di dapat dengan menggunakan rumus (2). Nilai RMSEP yang lebih rendah dan nilai R^2 yang lebih tinggi menyiratkan kinerja model yang baik (Hodson, 2022). Dimana y adalah data aktual, y' adalah data prediksi, n adalah banyak nya data, i adalah urutan data.

$$RMSEP = \sqrt{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2 / n_p} \tag{1}$$

$$R^2 = \left[\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \right]^2 \tag{2}$$

Hasil dan Pembahasan

Model Kalibrasi menggunakan metode PCR

Data pada gambar 2 dianalisis dengan fungsi *train()* dan metode PCR; namun, prediktor tersebut diturunkan menggunakan metode analisis komponen utama (PCA) sebelum pemodelan dengan metode PCR. Instalasi beberapa paket terlebih dahulu, termasuk paket *tidyverse*, paket *caret*, dan paket *pls*, sebelum menggunakan metode PCR dengan metode *train()*; untuk petunjuk tentang cara melakukannya, dapat dilihat pada Gambar 7.

```
1  "" {r}
2  Install.package("tidyverse")
3  Install.package("caret")
4  Install.package("pls")
5  ""
```

Gambar 7. Skrip kode R untuk instalasi package *tidyverse*, *caret* dan *pls*

Setelah melakukan instalasi *package* dilanjutkan dengan memakai *package* dengan cara perintah *library(tidyverse)*, *library(caret)*, *library(pls)*. Langkah selanjutnya adalah menggunakan metode PCR untuk membuat model kalibrasi data, kemudian menggunakan fungsi *train()* untuk membuat model dengan menggunakan variabel PAW sebagai variabel respon. Untuk mengevaluasi kinerja model, fungsi *train()* menggunakan validasi yang sama dengan "CV", yang merupakan singkatan dari k-fold cross validation. Secara *default*, menggunakan *k=10 fold*. Ketika *scale=TRUE* menjelaskan model bahwa setiap variabel prediktor harus diskalakan ke *mean 0* dan *standar deviasi 1*. Ini adalah penjelasan mengapa model tidak menyertakan variabel prediktor yang terlalu berpengaruh dalam model jika diukur dalam berbagai satuan, serta skrip kode R untuk membuat model seperti pada gambar 8. Pada gambar 9 menunjukkan hasil skrip kode R yang digunakan untuk melihat hasil model.

```
1  "" {r}
2  library(tidyverse)
3  library(caret)
4  library(pls)
5  model <- train (
6    PAW~., data = train.data, method = "pcr",
7    scale = TRUE,
8    trControl = trainControl("cv", number = 10),
9    tuneLength = 10
10 )
11 model
""
```

Gambar 8. Skrip kode R untuk instalasi package *tidyverse*, *caret* dan *pls*

```

10 samples
5 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 9, 9, 9, 9, 9, 9, ...
Resampling results across tuning parameters:

ncomp  RMSE      Rsquared  MAE
1      0.1427722  NaN       0.1427722
2      0.1463271  NaN       0.1463271
3      0.2744961  NaN       0.2744961
4      0.4045421  NaN       0.4045421

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was ncomp = 1.
    
```

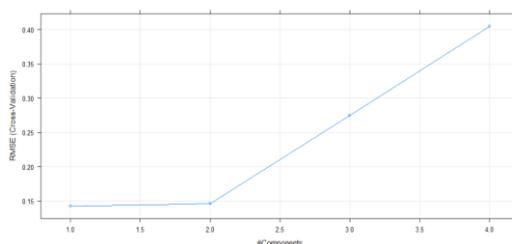
Gambar 9. Hasil Skrip kode R untuk melihat model.

Gambar 9 menunjukkan bahwa terdapat 10 *sampel* dalam model PCR, dengan lima variabel prediktor dan empat komponen, yang ditunjukkan oleh nilai RMSE masing-masing komponen. Visualisasi nilai RMSE untuk setiap komponen secara grafis menggunakan skrip kode R, seperti yang diilustrasikan pada gambar 10. Gambar 11 menunjukkan output nilai RMSE skrip kode R berupa tabel RMSE untuk setiap komponen menggunakan fungsi *plot()*.

```

1  ```{r}
2  # Plot model RMSE vs different values of components
3  plot(model)
4  ```
    
```

Gambar 10. Skrip kode R untuk melihat grafik nilai RMSE.



Gambar 11. Hasil skrip kode R untuk melihat grafik nilai RMSE

Pengujian model RMSE yang diperoleh dengan menggunakan *k-fold cross-validation* digambarkan pada Gambar 11. Terlihat bahwa ketika *intersep term* dimasukkan ke dalam model, RMSE pengujiannya adalah 14,27, namun ketika komponen utama pertama disertakan, RMSE *tes* adalah 14,63. Dengan demikian dapat dilihat bahwa peningkatan jumlah komponen utama sebenarnya meningkatkan RMSE pengujian. Akibatnya, tampaknya hanya menggunakan satu komponen utama dalam model akhir adalah pilihan terbaik. Setelah pembuatan model, diperlukan untuk mengujinya menggunakan *test.data* skrip kode R. Gambar 12 menunjukkan data yang digunakan untuk menghitung nilai prediksi menggunakan fungsi *predict()*. Gambar 13 menunjukkan hasil skrip kode R yang digunakan untuk menghitung nilai yang diharapkan dalam dataset.

```

1  ```{r}
2  # Make predictions
3  predictions <- model %>% predict(test.data)
4  test.data$predict_per <- predictions
5  test.data
6  ```
    
```

Gambar 12. Skrip kode R untuk mendapatkan nilai prediksi dari fungsi *predict()*

A tibble: 10 x 7

PAW <dbl>	SPEKTROFOTOMETRI ULTRAVIOLET <dbl>	Sensor_Cahaya <dbl>	Suhu <dbl>	Voltase KV <dbl>	Kipas <dbl>	predict_pcr <dbl>
1.85	0.5562	0.15	29.4	600	425	1.763838
1.20	0.3709	0.04	26.4	200	150	1.160529
1.95	0.6372	0.35	28.2	800	475	1.917879
2.00	0.6575	0.40	29.0	800	500	1.991632
1.32	0.4533	0.07	28.1	400	225	1.421395
2.10	0.6776	0.97	28.8	800	525	2.118375
1.92	0.5869	0.25	27.1	600	450	1.724368
1.69	0.5467	0.13	28.2	600	375	1.672811
1.55	0.4945	0.11	27.3	400	325	1.486308
2.30	0.6879	1.89	30.9	800	675	2.471247

10 rows

Gambar 13. Hasil skrip kode R untuk mendapatkan nilai prediksi dari data test.

Skrip kode R dapat dilihat pada gambar 14 setelah memperoleh data nilai yang diproyeksikan dalam kumpulan data dan menilai RMSE dan R^2 menggunakan *library(caret)*. Pada gambar 15 menunjukkan hasil skrip kode R yang digunakan untuk menghitung RMSE dan R^2 menggunakan *test.data*. Nilai uji RMSE adalah 0,09625, seperti yang ditunjukkan pada gambar 15. Ini adalah perbedaan antara nilai larutan PAW yang diprediksi dan nilai larutan PAW yang diamati pada *set* pengujian yang diamati. Sementara itu, uji R^2 sebesar 0,9304 menunjukkan bahwa semua faktor prediktor memiliki pengaruh gabungan terhadap larutan PAW sebesar 93,04 persen (variabel respons). Sedangkan variabel lain yang tidak dibahas dalam penelitian mempengaruhi sisanya sebesar 7,06 persen.

```

1   "" {r}
2   # Model performance metrics in PCR
3   data.frame(
4     RMSE = caret::RMSE(test.data$predict_pcr, test.data$PAW),
5     Rsquare = caret::R2(test.data$predict_pcr, test.data$PAW)
6   )
7   ""
    
```

Gambar 14. Skrip kode R untuk mendapatkan nilai RMSE dan R^2

Description: df [1 x 2]

	RMSE <dbl>	Rsquare <dbl>
1 row	0.09625571	0.9304699

Gambar 15. Hasil skrip kode R untuk mendapatkan nilai RMSE dan R^2

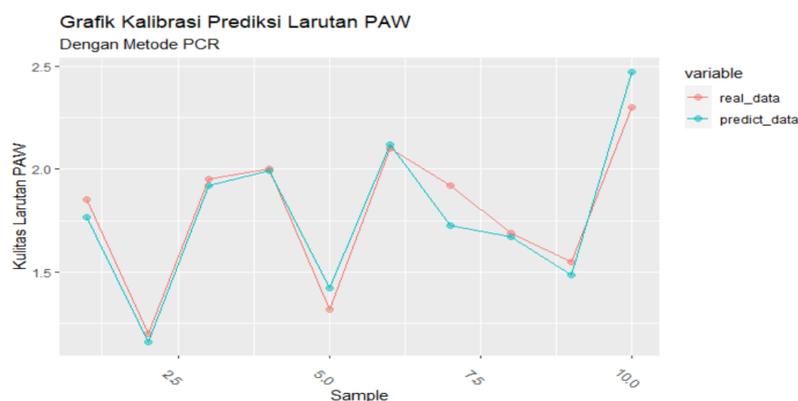
Skrip kode R pada Gambar 16 dapat digunakan untuk menampilkan data real dengan data prediktif menggunakan pendekatan PCR. Pada gambar 17 menunjukkan hasil skrip kode R untuk membuat grafik yang menggabungkan nilai data nyata dengan data prediksi. Data real dan data prediksi pada gambar 17 dapat dilihat pada grafik, dan dapat dikatakan bahwa model yang dibangun dengan pendekatan PCR hampir identik dengan data aslinya.

```

1  "" {r}
2  library(ggplot2)
3  library(reshape2)
4  x <- 1:10
5  real_data <- test.data$PAW
6  predict_data <- test.data$predict_pcr
7  df <- data.frame(x, real_data, predict_data)
8  # melt the data to a long format
9  df2 <- melt(data = df, id.vars = "x")
10 # plot, using the aesthetics argument 'colour'
11 ggplot(data = df2, aes(x = x, y = value, colour = variable)) +
12   geom_point(alpha = 1/2, size=2) +
13   geom_line() +
14   theme(legend.justification = "top") +
15   labs(title = "Grafik Kalibrasi Prediksi Larutan PAW ",
16         subtitle = "Dengan Metode PCR ",
17         y = "Kualitas Larutan PAW", x = "Sample") +
18   theme(axis.text.x = element_text(angle = -45))
19  ""

```

Grafik 16. Dengan menggunakan metode PCR, skrip kode R menghasilkan grafik untuk nilai data nyata dengan data prediksi.



Gambar 17. Hasil skrip kode R membuat grafik untuk nilai data real dengan data prediksi dengan metode PCR

Model kalibrasi menggunakan metode PLS-R

Fungsi *pls()* digunakan untuk melakukan analisis data menggunakan pendekatan PLS-R pada data larutan PAW pada gambar 1. Metode PLS-R dapat menyelesaikan masalah *multikolinearitas* pada data. *Multikolinearitas* adalah sebuah situasi pada data yang menunjukkan adanya korelasi kuat antara dua variabel bebas atau lebih dalam sebuah model regresi berganda (Daoud, 2017). Pendekatan PLS-R dapat mereduksi data menjadi *dataset* yang sempurna, dengan korelasi yang baik antara variabel prediktor dan respon, sehingga menghasilkan model kalibrasi data yang baik. Kemudian, dengan menggunakan teknik PLS-R, untuk membuat model menggunakan fungsi *train()* dengan metode *pls* dan variabel PAW sebagai variabel respon, dan terakhir, menggunakan fungsi *train()* dengan variabel PAW sebagai variabel respon untuk menghasilkan model.

Untuk menguji kinerja model, gunakan validasi = "CV", yang merupakan singkatan dari validasi silang *k-fold*. Secara menggunakan *default*, *k=10 fold*. Ketika menggunakan *scale=TRUE*, menjelaskan model bahwa setiap variabel prediktor harus diskalakan ke *mean 0* dan *standar deviasi 1*. Ini menjelaskan mengapa model tidak memiliki variabel prediktor yang terlalu berpengaruh dalam model jika diukur dalam berbagai unit. Model PLS-R menggunakan data train dibuat dengan skrip kode R seperti yang ditunjukkan pada gambar 17. Gambar 18 menunjukkan hasil skrip kode R yang digunakan untuk melihat model PLS-R.

```

1  "" {r}
2  model1 <- train(
3  PAW~, data = train.data, method = "pls",
4  scale = TRUE,
5  trControl = trainControl("cv", number = 10),
6  tuneLength = 10
7  )
8  model1
9  ""
    
```

Gambar 17. Skrip kode R membuat model1 dengan metode PLS-R

```

10 samples
5 predictor

No pre-processing
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 9, 9, 9, 9, 9, ...
Resampling results across tuning parameters:

ncomp RMSE      Rsquared MAE
1     0.1448042  NaN    0.1448042
2     0.1709893  NaN    0.1709893
3     0.3059224  NaN    0.3059224
4     0.4050681  NaN    0.4050681

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was ncomp = 1.
    
```

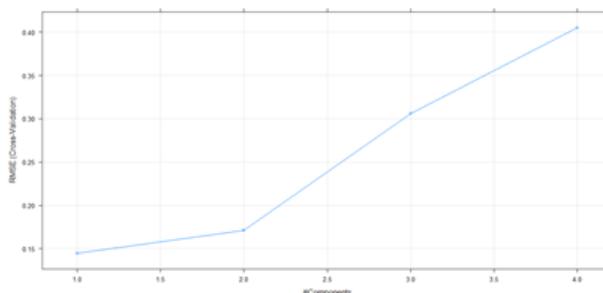
Gambar 18. Hasil skrip kode R membuat model1 dengan metode PLS-R

Pada model PLS-R terdapat 10 sampel dengan lima variabel prediktor dan empat komponen, yang ditunjukkan oleh nilai RMSE masing-masing komponen seperti yang terlihat pada gambar 18. Visualisasi nilai RMSE untuk setiap komponen secara grafis menggunakan skrip kode R, seperti diilustrasikan pada gambar 19. Gambar 20 menunjukkan hasil skrip kode R, yang merupakan grafik RMSE dengan nilai RMSE yang disajikan untuk setiap komponen menggunakan fungsi *plot()*.

```

1  "" {r}
2  plot(model1)
3  ""
    
```

Gambar 19. Skrip kode R membuat menampilkan model1 berupa nilai RMSE dalam bentuk grafik.



Gambar 20. Skrip kode R membuat menampilkan model1 berupa nilai RMSE pada setiap komponen.

Pada gambar 20 menjelaskan tentang pengujian model terhadap RMSE yang dihitung dengan validasi silang *k-fold*. Dapat dilihat bahwa jika menggunakan istilah *intersep* dalam model, RMSE pengujian adalah 0,1448, sedangkan jika menambahkan komponen utama pertama, *tes* RMSE meningkat menjadi 0,1709. Maka dapat dilihat bahwa menambahkan komponen utama tambahan sebenarnya mengarah pada peningkatan RMSE pengujian. Dengan demikian, tampaknya optimal hanya menggunakan satu komponen utama dalam model akhir. Pengujian model RMSE, yang diperoleh dengan validasi silang *k-fold*, digambarkan pada gambar 20. RMSE pengujian adalah 0,1448, seperti yang dapat diamati, ketika istilah *intersep* dimasukkan dalam model, namun RMSE dari *tes* naik ke 0,1709 ketika komponen utama pertama disertakan. Dengan demikian dapat dilihat bahwa peningkatan jumlah komponen utama sebenarnya meningkatkan RMSE pengujian.

Setelah membuat model maka perlu untuk mencoba model dengan *test.data* skrip kode R untuk mendapatkan nilai prediksi dari fungsi *predict()* dapat dilihat pada gambar 21. Akibatnya, tampaknya hanya menggunakan satu komponen utama dalam model akhir adalah pilihan terbaik. Setelah

pembuatan model, diperlukan untuk mengujinya. Gambar 21 menunjukkan data dalam kode R untuk menghitung nilai prediksi dari fungsi *predict()*. Gambar 22 menunjukkan hasil skrip kode R yang digunakan untuk menghitung nilai yang diharapkan dalam *dataset*. Gambar 23 menunjukkan hasil skrip kode R yang digunakan untuk menghitung RMSE dan R^2 menggunakan *test.data*.

```

1  "" {r}
2  # Make predictions
3  predictions <- model1 %>% predict(test.data)
4  test.data$predict_pcr <- predictions
5  test.data
6  ""
    
```

Gambar 21. Skrip kode R membuat nilai prediksi dari test data

PAW	SPECTROFOTOMETRI ULTRAVIOLET	Sensor_Cahaya	Suhu	Voltage KV	Kipas	predict_pcr	predict_plsr
1.85	0.5562	0.15	29.4	800	425	1.762838	1.759487
1.20	0.3769	0.04	28.4	200	150	1.865259	1.862281
1.95	0.6372	0.25	28.2	800	475	1.977879	1.978086
2.00	0.6575	0.40	29.0	800	500	1.991632	1.990647
1.32	0.4533	0.07	28.1	400	225	1.421395	1.418930
2.10	0.6776	0.97	28.8	800	525	2.118375	2.125159
1.92	0.5969	0.25	27.3	800	450	1.724068	1.728537
1.69	0.5487	0.13	28.2	600	375	1.672811	1.670707
1.55	0.4945	0.11	27.3	400	325	1.488208	1.489568
2.30	0.8879	1.89	30.9	800	675	2.471247	2.483778

Gambar 22. Hasil skrip kode R membuat nilai prediksi dari test data.

```

1  "" {r}
2  # Model performance metrics in PLS-R
3  data.frame(
4  RMSE = caret::RMSE(test.data$predict_plsr, test.data$PAW),
5  Rsquare = caret::R2(test.data$predict_plsr, test.data$PAW)
6  )
7  ""
    
```

Gambar 23. Skrip kode R membuat nilai RMSE dan R^2 dari model PLS-R

Gambar 24 menunjukkan hasil skrip kode R yang digunakan untuk menghasilkan nilai RMSE dan R^2 dari model berdasarkan data pengujian. Hasil pengujian RMSE adalah 0,0987, seperti yang ditunjukkan pada gambar 24. Ini adalah perbedaan antara nilai larutan PAW yang diprediksi dan nilai larutan PAW yang diamati pada set pengujian yang diamati. Sementara itu, uji R^2 sebesar 0,9244 menunjukkan bahwa semua faktor prediktor memiliki efek gabungan terhadap larutan PAW sebesar 92,44 persen (variabel respons).

	RMSE	Rsquare
	<dbl>	<dbl>
1 row	0.09873341	0.9284436

Gambar 24. Hasil skrip kode R untuk membuat nilai RMSE dan R^2 dari model PLS-R

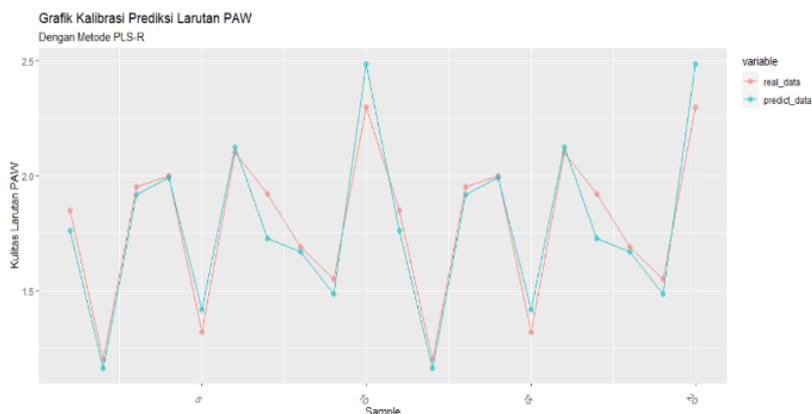
Sedangkan variabel lain yang tidak dibahas dalam penelitian mempengaruhi sisanya sebesar 7,56 persen. Skrip kode R pada gambar 25 dapat digunakan untuk menampilkan grafik data nyata dan data prediksi. Gambar 26 menunjukkan hasil skrip kode R untuk membuat grafik data aktual dengan data prediksi dari model PLS-R. Data asli dan data prediksi pada gambar 26 dapat dilihat pada grafik, dan model yang dibangun menggunakan pendekatan PLS-R hampir identik dengan data sebenarnya.

```

1  ```{r}
2  library(ggplot2)
3  library(reshape2)
4  x <- 1:20
5  real_data <- test.data$PAW
6  predict_data <- test.data$predict_plsr
7  df <- data.frame(x, real_data, predict_data)
8  # melt the data to a long format
9  df2 <- melt(data = df, id.vars = "x")
10 # plot, using the aesthetics argument 'colour'
11 ggplot(data = df2, aes(x = x, y = value, colour = variable))+
12   geom_point(alpha = 1/2,size=2) +
13   geom_line()+
14   theme(legend.justification = "top") +
15   labs(title = "Grafik Kalibrasi Prediksi Larutan PAW ",
16        subtitle = "Dengan Metode PLS-R ",
17        y = "Kulitas Larutan PAW", x = "Sample") +
18   theme(axis.text.x = element_text(angle = -45))
19  ```

```

Gambar 25. Skrip kode R membuat grafik data real dengan data prediksi dari model PLS-R



Gambar 26. Skrip kode R membuat grafik data real dengan data prediksi dari model PLS-R

Perbandingan model kalibrasi menggunakan metode PCR dan metode PLS-R

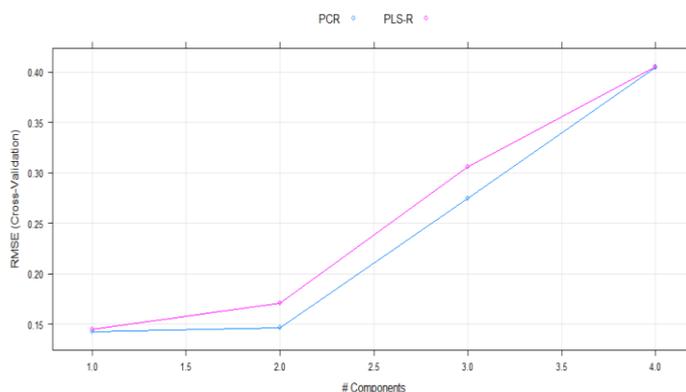
Untuk membandingkan nilai RMSE masing-masing komponen dari kedua pendekatan dengan grafik setelah mendapatkan nilai RMSE dari metode PCR dan PLS-R, skrip kode R seperti yang terlihat pada gambar 27. Hasil skrip kode R untuk nilai RMSE dari metode PCR dan PLS-R dapat dilihat pada gambar 28.

```

1  ```{r}
2  # Save PLS model tuning information to plsResamples
3  plsResamples <- model1$results
4  plsResamples$Model <- "PLS-R"
5  # Save PCR model tuning information to plsResamples
6  pcrResamples <- model2$results
7  pcrResamples$Model <- "PCR"
8  # Combine both output for plotting
9  plsPlotData <- rbind(plsResamples, pcrResamples)
10 # Leverage xyplot() function from lattice library
11 xyplot(RMSE ~ ncomp,
12        data = plsPlotData,
13        xlab = "# Components",
14        ylab = "RMSE (Cross-Validation)",
15        auto.key = list(columns = 2),
16        groups = Model,
17        type = c("o", "g"))
18  ```

```

Gambar 27. Skrip kode R membuat grafik data RMSE dari model menggunakan metode PCR dan metode PLS-R



Gambar 28. Hasil skrip kode R membuat grafik data RMSE dari model menggunakan metode PCR dan metode PLS-R

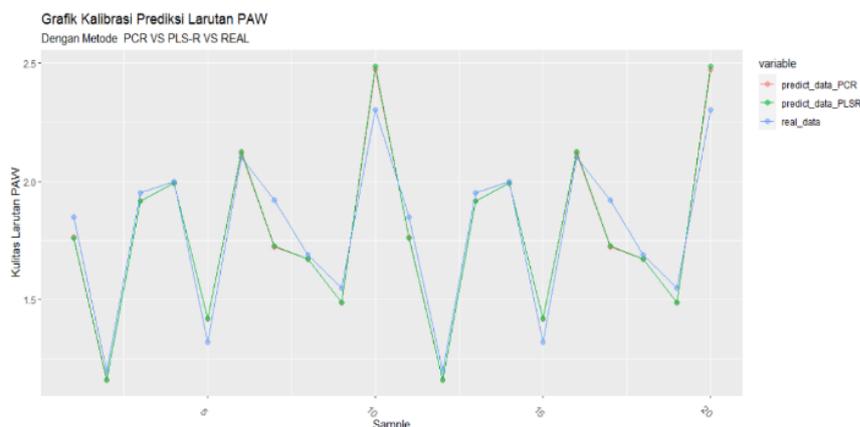
Sedangkan untuk membandingkan tiga data yaitu data real, data prediksi dari metode PCR dan data prediksi dari metode PLS-R dapat menggunakan skrip program R pada gambar 29. Hasil skrip program R data real, data prediksi dari metode PCR dan data prediksi dari metode PLS-R bentuk grafik dapat dilihat pada gambar 30.

```

1  ```{r}
2  library(ggplot2)
3  library(reshape2)
4  x <- 1:20
5  real_data <- test.data$PAW
6  predict_data_PCR <- test.data$predict_per
7  predict_data_PLSR <- test.data$predict_plsr
8  df <- data.frame(x, predict_data_PCR, predict_data_PLSR, real_data)
9  # melt the data to a long format
10 df2 <- melt(data = df, id.vars = "x")
11 # plot, using the aesthetics argument 'colour'
12 ggplot(data = df2, aes(x = x, y = value, colour = variable)) +
13   geom_point(alpha = 1/2, size=2) +
14   geom_line()+
15   theme(legend.justification = "top") +
16   labs(title = "Grafik Kalibrasi Prediksi Larutan PAW ",
17        subtitle = "Dengan Metode PCR VS PLS-R VS REAL ",
18        y = "Kulitas Larutan PAW", x = "Sample") +
19   theme(axis.text.x = element_text(angle = -45))
20  ```

```

Gambar 29. Skrip kode R membuat grafik data real, data prediksi PCR dan data PLS-R



Gambar 30. Hasil skrip kode R membuat grafik data real, data prediksi PCR dan data prediksi PLS-R

Pada gambar 30, grafik ketiga data yaitu data real, data prediksi dapat dilihat bahwa model yang dihasilkan menggunakan metode PCR dan data prediksi dapat diamati bahwa model yang dibangun menggunakan metode PLS-R hampir identik. Perbandingan kedua metode tersebut dapat dinyatakan bahwa metode PCR lebih efisien daripada metode PLS-R untuk membangun model kalibrasi untuk

prediksi larutan PAW dari reaktor *plasma*. Pada penelitian ini menunjukkan bahwa PCR mengungguli PLS-R dalam akurasi prediktif karena data pengamatan adalah data yang sederhana dan data pengamatan memiliki dimensi rendah dengan *multikolinearitas* sehingga berbeda dengan penelitian sebelumnya yang menyatakan bahwa PLS-R adalah lebih unggul (Araújo et al., 2014 ; Gholizadeh et al., 2017 ; Hashem et al., 2022). Keuntungan penerapan metode PCR pada penelitian adalah metode ini hanya mengurangi gangguan tetapi masih mempertahankan varians data (Malik, 2015 ; He et al., 2005).

Kesimpulan

Metode PCR dan PLS-R dapat digunakan untuk membangun model kalibrasi. Ketika kedua metode tersebut digunakan untuk model kalibrasi data dari reaktor PAW, maka dapat dinyatakan bahwa metode PCR lebih efisien daripada metode PLS-R untuk membangun model kalibrasi untuk prediksi larutan PAW dari reaktor *plasma*. Hal ini ditunjukkan bahwa nilai R^2 sebesar 93,04699% pada model kalibrasi menggunakan metode PCR adalah lebih besar yang menandakan nilai varian data dapat diterima oleh model kalibrasi dibanding dengan nilai R^2 sebesar 92,84436% pada metode PLS-R, kemudian nilai RMSEP sebesar 0,09625571 pada menggunakan metode PCR lebih kecil yang menandakan nilai *error* statistik pada perhitungan dapat diterima dibanding nilai RMSEP sebesar 0,09873341 pada metode PLS-R.

Ucapan Terima Kasih

Penelitian ini didanai oleh Lembaga Penelitian dan Pengabdian Masyarakat (LP2M) UIN Maulana Malik Ibrahim Malang tahun anggaran 2021.

Daftar Pustaka

- Ahmed, M., & Kim, D. R. (2018). pcr: An R package for quality assessment, analysis and testing of qPCR data. *PeerJ*, 6, e4473. <https://doi.org/10.7717/peerj.4473>
- Araújo, S. R., Wetterlind, J., Demattê, J. A. M., & Stenberg, B. (2014). Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. *European Journal of Soil Science*, 65(5), 718–729. <https://doi.org/10.1111/ejss.12165>
- Arsham, H., & Lovric, M. (2011). Bartlett's Test. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 87–88). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_132
- Baracchini, T., Hummel, S., Verlaan, M., Cimatoribus, A., Wüest, A., & Bouffard, D. (2020). An automated calibration framework and open source tools for 3d lake hydrodynamic models. *Environmental Modelling & Software*, 134, 104787. <https://doi.org/10.1016/j.envsoft.2020.104787>
- Buker, E., Dinç, E., & Yasar, N. (2022). Application of pcr and pls tools for the simultaneous quantification of praziquantel and ivermectin binary mixtures in veterinary tablets. *Düzce Üniversitesi Bilim Ve Teknoloji Dergisi*, 10(1), 385-397. <https://doi.org/10.29130/dubited.1051389>
- Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, 949, 012009. <https://doi.org/10.1088/1742-6596/949/1/012009>
- Da Silva, R. M., Dantas, J. C., Beltrão, J. D. A., & Santos, C. A. G. (2018). Hydrological simulation in a tropical humid basin in the Cerrado biome using the SWAT model. *Hydrology Research*, 49(3), 908–923. <https://doi.org/10.2166/nh.2018.222>

- Gholizadeh, A., Carmon, N., Klement, A., Ben-Dor, E., & Borůvka, L. (2017). Agricultural Soil Spectral Response and Properties Assessment: Effects of Measurement Protocol and Data Mining Technique. *Remote Sensing*, 9(10), 1078. <https://doi.org/10.3390/rs9101078>
- Hashem, M., Islam, M., Hossain, M., Alam, A., & Khan, M. (2022). Prediction of chevon quality through near infrared spectroscopy and multivariate analyses. *Meat Research*, 2(6). <https://doi.org/10.55002/mr.2.6.37>
- He, Y., Li, X., Hernandez, A., & Garcia, A. (2005). Vis/NIR spectroscopy technique for determination quality attributes of tomato fruit (L. Zhang, J. Zhang, & M. Liao, Eds.; p. 60431Y). <https://doi.org/10.1117/12.654948>
- Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. <https://doi.org/10.5194/gmd-2022-64>
- M., Y., Rao, V. G., Devangad, P., D'Souza, J. S., & Chidangil, S. (2020). A chemometric study combined with spectroscopy for the quantification of secondary structure of flagellar-associated protein 174 (FAP174). *Journal of Chemometrics*, 34(5), e3221. <https://doi.org/10.1002/cem.3221>
- Mai, J. (2023). Ten strategies towards successful calibration of environmental models. *Journal of Hydrology*, 620, 129414. <https://doi.org/10.1016/j.jhydrol.2023.129414>
- Malik, B. A. (2015). Determination of glucose concentration from near infrared spectra using least square support vector machine. 2015 International Conference on Industrial Instrumentation and Control (ICIC), 475–478. <https://doi.org/10.1109/IIC.2015.7150789>
- Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software*, 18(2). <https://doi.org/10.18637/jss.v018.i02>
- Pandey, A. N., Pathak, P. C., & Singh, J. S. (1983). Water, sediment and nutrient movement in forested and non-forested catchments in Kumaun Himalaya. *Forest Ecology and Management*, 7(1), 19–29. [https://doi.org/10.1016/0378-1127\(83\)90054-3](https://doi.org/10.1016/0378-1127(83)90054-3)
- Pianosi, F., & Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrological Processes*, 30(22), 3991–4003. <https://doi.org/10.1002/hyp.10968>
- Rácz, A., Vass, A., Héberger, K., & Fodor, M. (2015). Quantitative determination of coenzyme Q10 from dietary supplements by FT-NIR spectroscopy and statistical analysis. *Analytical and Bioanalytical Chemistry*, 407(10), 2887–2898. <https://doi.org/10.1007/s00216-015-8506-8>
- Rasmussen, P. M., Smith, A. F., Sakadžić, S., Boas, D. A., Pries, A. R., Secomb, T. W., & Østergaard, L. (2017). Model-based inference from microvascular measurements: Combining experimental measurements and model predictions using a Bayesian probabilistic approach. *Microcirculation*, 24(4), e12343. <https://doi.org/10.1111/micc.12343>
- Saleh, D., Wang, G., Müller, B., Rischawy, F., Kluters, S., Studts, J., & Hubbuch, J. (2020). Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications. *Biotechnology Progress*, 36(4), e2984. <https://doi.org/10.1002/btpr.2984>
- Singh, I., Juneja, P., Kaur, B., & Kumar, P. (2013). Pharmaceutical Applications of Chemometric Techniques. *ISRN Analytical Chemistry*, 2013, 1–13. <https://doi.org/10.1155/2013/795178>
- Sung, C., & Tuo, R. (2024). A review on computer model calibration. *WIREs Computational Statistics*, 16(1), e1645. <https://doi.org/10.1002/wics.1645>
- Tan, C., Li, M., & Qin, X. (2008). Random Subspace Regression Ensemble for Near-Infrared Spectroscopic Calibration of Tobacco Samples. *Analytical Sciences*, 24(5), 647–653. <https://doi.org/10.2116/analsci.24.647>
- Thirumdas, R., Kothakota, A., Annapure, U., Siliveru, K., Blundell, R., Gatt, R., & Valdramidis, V. P. (2018). Plasma activated water (PAW): Chemistry, physico-chemical properties, applications in food and agriculture. *Trends in Food Science & Technology*, 77, 21–31. <https://doi.org/10.1016/j.tifs.2018.05.007>

- Tuo, Y., Marcolini, G., Disse, M., & Chiogna, G. (2018). A multi-objective approach to improve SWAT model calibration in alpine catchments. *Journal of Hydrology*, 559, 347–360. <https://doi.org/10.1016/j.jhydrol.2018.02.055>
- Van Meerveld, H. J. I., Vis, M. J. P., & Seibert, J. (2017). Information content of stream level class data for hydrological model calibration. *Hydrology and Earth System Sciences*, 21(9), 4895–4905. <https://doi.org/10.5194/hess-21-4895-2017>
- Waldron, A., Ulusoy, A.-J., Pecci, F., & Stoianov, I. (2022). Principal Component Based Sampling for the Continuous Maintenance of Hydraulic Models. *Water Research*, 222, 118905. <https://doi.org/10.1016/j.watres.2022.118905>
- Wijewardane, N. K., Ge, Y., Sanderman, J., & Ferguson, R. (2021). Fine grinding is needed to maintain the high accuracy of mid-infrared diffuse reflectance spectroscopy for soil property estimation. *Soil Science Society of America Journal*, 85(2), 263–272. <https://doi.org/10.1002/saj2.20194>
- Zhao, Y., Patange, A., Sun, D., & Tiwari, B. (2020). Plasma-activated water: Physicochemical properties, microbial inactivation mechanisms, factors influencing antimicrobial effectiveness, and applications in the food industry. *Comprehensive Reviews in Food Science and Food Safety*, 19(6), 3951–3979. <https://doi.org/10.1111/1541-4337.12644>