

PROPOSAL DISERTASI  
**MODEL PENDUGA CURAH HUJAN ZONA MUSIM  
MENGUNAKAN PREDIKTOR SIMULTAN  
DAN HYBRID ARTIFICIAL NEURAL NETWORKS  
(Studi Kasus : Wilayah Propinsi Jawa Timur)**

Ahmad Saikhu  
5115301007

DOSEN PEMBIMBING  
Dr. Agus Zainal Arifin, S.Kom, M.Kom  
Dr. Chastine Fatichah, S.Kom, M.Kom

PROGRAM DOKTOR  
PROGRAM STUDI ILMU KOMPUTER  
JURUSAN TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017



## LEMBAR PENGESAHAN

Judul : Model Penduga Curah Hujan Zona Musim Menggunakan Prediktor  
Simultan dan hybrid Artificial Neural Networks  
(Studi Kasus Wilayah Propinsi Jawa Timur)

Oleh : Ahmad Saikhu

NRP. : 5115301007

### Telah diseminarkan pada:

Hari : Kamis

Tanggal : 5 Oktober 2017

Tempat : Ruang Sidang Departemen Informatika

### Mengetahui/menyetujui:

1. Dr. Agus Zainal Arifin, S.Kom, M.Kom.  
NIP. 19720809 199512 1 001  
.....  
(Pembimbing 1)
2. Dr. Chastine Fatichah, S.Kom, M.Kom  
NIP. 19751220 200112 2 002  
.....  
(Pembimbing 2)
3. Drs. Edi Winarko, M.Sc, Ph.D  
NIP. 19630223 1987 03 1 002  
.....  
(Penguji Eksternal)
4. Dr. Suhartono, S.Si, M.Sc  
NIP. 19710929 1995 12 1 001  
.....  
(Penguji 1)
5. Dr. Darlis Herumurti, S.Kom, M. Kom  
NIP. 19771217 2003 12 1 001  
.....  
(Penguji 2)



# MODEL PENDUGA CURAH HUJAN ZONA MUSIM MENGUNAKAN PREDIKTOR SIMULTAN DAN HYBRID ARTIFICIAL NEURAL NETWORKS

Nama mahasiswa : Ahmad Saikhu  
NRP : 5115301007  
Pembimbing I : Dr. Agus Zainal Arifin, S.Kom. M.Kom  
Pembimbing II : Dr. Chastine Fatichah, S.Kom, M.Kom

## ABSTRAK

Cuaca, iklim dan musim senantiasa akrab dalam kehidupan manusia sehari-hari. Salah satu komponen penting dalam cuaca dan iklim adalah curah hujan. Curah hujan di suatu wilayah diakibatkan oleh interaksi berbagai variabel cuaca dan iklim global. Curah hujan merupakan sistem yang kompleks karena hubungan antar variabel lokal bersifat simultan dan dipengaruhi waktu, wilayah dan kondisi iklim global yang dinamis. Oleh karenanya, hubungan antar variabel tidak hanya bersifat sesaat tetapi terkait dengan kondisi wilayah sekitar dan keadaan cuaca/iklim pada periode sebelumnya.

Pemodelan curah hujan di suatu wilayah cukup penting untuk perencanaan sumber daya air secara optimal, pertanian, infrastruktur, pertambangan dan merupakan penelitian penting dalam bidang hidrologi. Berbagai penelitian telah dilakukan untuk pemodelan curah hujan, baik pendekatan model teoritis/konseptual maupun pendekatan data empiris/histori. Pemodelan konseptual bertumpu pada interaksi dan proses antara dinamika atmosfer, suhu permukaan laut, dan posisi matahari. Oleh karenanya model ini memiliki kompleksitas tinggi dan memerlukan komputasi yang mahal. Model empiris dipilih untuk menemukan hubungan antara data input dan output ketika pengguna tidak memahami secara lengkap hubungan faktor-faktor fisik yang terjadi dan lebih berfokus pada akurasi.

Berdasarkan studi, curah hujan di Indonesia dipengaruhi oleh berbagai faktor lokal (cuaca) dan iklim regional (global). Oleh karenanya penentuan prediktor yang relevan adalah penting untuk efisiensi model penduga curah hujan. Meskipun beberapa penelitian pemodelan iklim di suatu wilayah/negara memiliki akurasi tinggi, belum tentu model tersebut sesuai untuk wilayah lainnya dikarenakan berbagai faktor. Hal ini menyebabkan bahwa pemodelan untuk iklim dan prediksi hujan menjadi khas untuk suatu wilayah. Model iklim dan curah hujan wilayah tropis yang mampu mensimulasikan dan memprediksi curah hujan di Indonesia secara efektif belum ditemukan.

Salah satu pendekatan model empiris adalah metode ARIMA yang memiliki kelemahan kurang sesuai untuk data non-linier sehingga akurasinya rendah. Untuk mengatasi hal ini, salah satu pilihan adalah menggunakan model ANN yang juga memiliki kelemahan bahwa untuk pola linier, tidak menjamin akurasi yang optimal dan bergantung pada karakteristik data.

Untuk mengatasi sejumlah masalah di atas, adalah dengan mengidentifikasi sebanyak mungkin variabel prediktor dengan mempertimbangkan faktor waktu dan lokasi stasiun cuaca (spatio-temporal), fitur lokal/global dan pola hubungan linier/non linier. Oleh karenanya dalam penelitian ini diusulkan suatu pendekatan baru model penduga curah hujan wilayah Zona Musim di Indonesia melalui kombinasi proses ekstraksi dan seleksi variabel prediktor serta penggabungan (*hybrid*) model *spatio-temporal* ARX dan ANN. Kinerja model diukur dengan RMSE, MAE,  $R^2$ ,  $r$  dan *corrected* AIC.

Dengan pendekatan baru, baik dari segi prediktor yang simultan maupun model hybrid, prediksi curah hujan diharapkan hasilnya lebih akurat sehingga dapat digunakan sebagai informasi perencanaan di sejumlah bidang. Pemilihan prediktor dengan hanya melibatkan fitur relevan (signifikan) menjadikan proses komputasi lebih efisien.

## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
ABSTRAK.....	iii
DAFTAR ISI.....	v
DAFTAR TABEL.....	vii
DAFTAR GAMBAR .....	ix
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	4
1.3 Batasan Masalah .....	5
1.4 Tujuan dan Manfaat Penelitian .....	5
1.5 Kontribusi Penelitian .....	6
BAB 2 KAJIAN PUSTAKA DAN DASAR TEORI .....	7
2.1 Kajian Pustaka .....	7
2.1.1 Penelitian terkait model penduga curah hujan .....	7
2.1.2 Penelitian terkait pemilihan variabel prediktor .....	9
2.1.3 Prediktor tenggang waktu, input relevan dan tidak redundant.....	10
2.1.4 Penelitian spatio-temporal dan model <i>hybrid</i> .....	11
2.2 Dasar Teori.....	13
2.2.1 Cuaca dan iklim .....	13
2.2.2 Data runtun waktu multivariat .....	16
2.2.3 Ekstraksi fitur tenggang waktu ( <i>time-lags</i> ).....	19
2.2.4 Seleksi fitur .....	21
2.2.5 Model ARMA, ARIMA dan ARIMAX.....	23
2.2.6 Model artificial neural networks (ANN).....	25
2.2.7 Model <i>hybrid</i> ARIMA dan ANN .....	26
2.2.8 Data sintetik dan simulasi .....	27
2.2.9 Ukuran kinerja model.....	28
BAB 3 METODE PENELITIAN .....	31
3.1 Tahap Interpolasi dan Agregasi .....	33
3.2 Tahap Ekstraksi dan Seleksi Fitur.....	34
3.2.1 Ekstraksi fitur tenggang waktu ( <i>time-lags</i> ).....	34
3.2.2 Tahap seleksi fitur relevan dengan PCSUF .....	35
3.3 Pemodelan <i>Hybrid</i> ST-ARIMAX dan ANN.....	37

3.4 Evaluasi Kinerja .....	41
3.4.1 Evaluasi model dengan data sintetik .....	41
3.4.2 Evaluasi model dengan data riil.....	42
3.5 Hasil-Hasil Penelitian Awal .....	43
DAFTAR PUSTAKA .....	45



## **DAFTAR TABEL**

Tabel 3.1 Skenario evaluasi untuk data sintetik.....	42
Tabel 3.2 Skenario evaluasi untuk model penduga curah hujan.....	43
Tabel 3.3 karakteristik variabel curah hujan, prediktor lokal dan global .....	44



## DAFTAR GAMBAR

Gambar 2.1 Unsur-unsur cuaca dan iklim [116].....	14
Gambar 2.2 Zona Musim di Jawa Timur (sumber: web BMKG).....	15
Gambar 2.3 Zona ENSO di samudra pasifik [119].....	16
Gambar 2.4 Contoh data runtun waktu meteorologi [121] .....	17
Gambar 2.5 Plot data runtun waktu curah hujan [121].....	17
Gambar 2.6 Data runtun waktu non-stasioner [121].....	18
Gambar 2.7 ACF dan PACF dari curah hujan di Surabaya [121].....	18
Gambar 2.8 <i>Correlogram</i> CCF antara curah hujan dan suhu minimum [124].....	19
Gambar 2.9 DPCCA antara curah hujan, Nino3 dan PDO [91] .....	21
Gambar 2.10 Fitur prediktor yang relevan dan tidak relevan [98].....	23
Gambar 2.11 Flow-chart pemodelan ARIMA [121].....	24
Gambar 3.1 Diagram alir model penduga curah hujan .....	32
Gambar 3.2 Flow-chart untuk ekstraksi fitur <i>time-lags</i> dengan DPCCA [91] .....	34
Gambar 3.3 Seleksi fitur relevan dan hapus fitur redundan [124] .....	36
Gambar 3.4 Pra-proses, hybrid ST-ARX dan ANN .....	38
Gambar 3.5 Model <i>hybrid</i> STARX-ANN.....	39
Gambar 3.6 D etil proses pemodelan ST-ARX .....	39
Gambar 3.7 Spatio-temporal ANN .....	40



# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Cuaca, iklim dan musim senantiasa akrab dalam kehidupan manusia sehari-hari. Gejala alam tersebut berpengaruh kepada banyak sektor kehidupan, diantaranya pertanian, kehutanan, kemaritiman, perikanan, kesehatan, transportasi, pariwisata, pertambangan, energi dan infrastruktur. Dalam kondisi yang ekstrim, cuaca dan iklim bahkan dapat menyebabkan terjadinya bencana, diantaranya wabah penyakit, banjir, angin kencang, tanah longsor, kekeringan dan kebakaran hutan/lahan. Oleh karenanya, informasi, analisis dan pengetahuan terhadap kondisi cuaca dan iklim serta pola perubahan iklim sangat penting untuk dipelajari agar manusia dapat beradaptasi dengan perubahan yang dinamis, menghindarkan atau mengurangi terjadinya resiko akibat bencana dan melakukan tindakan mitigasi.

Dalam mengenali pola cuaca dan iklim dilakukan pengukuran terhadap suhu udara, kelembaban relatif, tekanan udara, penguapan air, kecepatan angin, suhu permukaan laut, lamanya penyinaran matahari dan curah hujan [1]. Nilai variabel tersebut berbeda bergantung pada sejumlah faktor, yaitu ketinggian tempat, garis lintang, tekanan, arus laut dan permukaan tanah. Batas nilai variabel untuk keadaan yang stabil telah ditentukan, namun dapat terjadi penyimpangan yang mengakibatkan masalah tertentu, misalnya hujan dan suhu ekstrim.

Salah satu komponen penting dalam cuaca dan iklim adalah curah hujan. Curah hujan di suatu wilayah diakibatkan oleh interaksi berbagai variabel cuaca dan iklim global. Curah hujan merupakan sistem yang kompleks di mana hubungan antar variabel bersifat simultan dan dipengaruhi waktu, wilayah dan kondisi iklim global yang dinamis. Oleh karenanya, hubungan antar variabel tidak hanya bersifat sesaat tetapi terkait dengan kondisi wilayah sekitar dan observasi cuaca/iklim pada periode sebelumnya.

Pemodelan curah hujan di suatu wilayah cukup penting untuk perencanaan sumber daya air secara optimal, perkiraan cadangan air di musim kering/kemarau, kegiatan pertambangan dan merupakan penelitian penting dalam bidang hidrologi.

Berbagai penelitian telah dilakukan untuk pemodelan curah hujan, baik pendekatan model teoritis/konseptual maupun pendekatan data empiris/histori. Pemodelan konseptual bertumpu pada interaksi/proses fisis dinamika atmosfer, suhu permukaan laut, dan posisi matahari [2].

Model empiris efektif untuk menemukan hubungan antara data input dan output ketika pengguna tidak memahami secara lengkap hubungan faktor-faktor fisik yang terjadi [3], namun lebih mementingkan nilai akurasi prediksi curah hujan. Di samping itu, pendekatan konseptual memerlukan proses komputasi yang mahal, kompleksitas tinggi dan padat peralatan *scientific* yang dalam beberapa tahun memerlukan pembaharuan peralatan [4].

Saat ini pemodelan empiris yang banyak digunakan untuk prediksi curah hujan adalah pendekatan statistik dan *data mining/machine learning*. Prediksi curah hujan dengan metode statistik adalah menggunakan pendekatan *time-series*, antara lain dengan *Autoregressive Integrated Moving Average* (ARIMA) dan *Seasonal ARIMA with Exogenous Input* (SARIMAX). Pendekatan ini memiliki sejumlah keterbatasan diantaranya mempersyaratkan data bersifat stasioner dan pola hubungan yang linier [3]. Sedangkan dalam permasalahan hidrologi dan lingkungan, pola data lebih banyak bersifat non linier [5], [6].

Untuk memperbaiki kelemahan dari model statistik, telah dikembangkan berbagai model prediksi curah hujan yang berbasis pada kecerdasan buatan, baik menggunakan metode *machine-learning* tunggal maupun gabungan (*hybrid*) dari metode statistik dan *machine-learning*. Metode *machine-learning* untuk prediksi curah hujan harian/bulanan diantaranya adalah *Bayesian Network* untuk analisis cuaca di wilayah India dan Canada [7], digunakannya *k-NN* dan *Mutual Information* (MI) untuk prediksi curah hujan dari data meteorologi di wilayah Dalian, China [8].

Selanjutnya, Ramana dkk. [9] menggunakan metode *Wavelet Neural Network* untuk data curah hujan bulanan di wilayah India. *Support Vector Machine* juga digunakan untuk prediksi curah hujan harian [10], [11]. Pendekatan *Neuro-Fuzzy* antara lain dilakukan oleh Hashim dkk. [12] untuk prediksi curah hujan bulanan di wilayah Patna, India. Pendekatan *Machine Learning* memiliki sifat dapat mengakomodasi pola non linier yang cukup baik.

Pendekatan model *hybrid* dan *ensemble* antara lain adalah dengan menggabungkan metode *Self Organized Map* (SOM) dan *Non-linear Autoregressive with Exogenous Inputs* (NARX) [13], ataupun menggabungkan *Multiple Linear Regression* (MLR) dan *Extreme Learning Machine* (ELM) untuk memprediksi curah hujan bulanan dengan prediktor anomali suhu permukaan laut [14]. Model hybrid DARIMA-NARX digunakan oleh Banihabib dkk. [15] untuk prediksi curah hujan harian dengan prediktor *North Atlantic Oscillation* (NAO). Penggabungan *Artificial Neural Networks* (ANN), MI dan *Particle Swarm Optimization* (PSO) digunakan untuk prediksi curah hujan bulanan di wilayah Australia [16]. Asadi dkk. [3] menggabungkan ANN dengan *Genetic Algorithm* (GA) dan algoritma *Levenverg-Marquardt* (LM) untuk prediksi curah hujan harian di wilayah India.

Pemodelan curah hujan dengan studi kasus wilayah Indonesia telah dilakukan dengan pendekatan ARIMA maupun *machine-learning*. Sejumlah penelitian prediksi curah hujan di wilayah Indonesia antara lain dengan studi kasus di Semarang, Tenggarong, Jakarta, Bandung dan Malang [17]–[25]. Metode yang digunakan antara lain ARIMA, ANN, ANFIS, sistem *Fuzzy* dan *Support Vector Regression*.

Berdasarkan sejumlah studi, dinyatakan bahwa curah hujan di Indonesia dipengaruhi oleh berbagai faktor lokal (cuaca) dan iklim regional/global [26]–[30]. Oleh karenanya penentuan variabel prediktor yang relevan adalah penting untuk efisiensi model dengan mempertimbangkan pola hubungan antar variabel yang kompleks.

Meskipun beberapa penelitian pemodelan iklim di suatu wilayah/negara memiliki akurasi yang tinggi, belum tentu model tersebut sesuai untuk wilayah lainnya dikarenakan perbedaan iklim, posisi geografis, topografi, dan proporsi daratan/maritim [31]. Sebagai contoh, model iklim untuk wilayah/negara lintang tinggi (sub tropis), akan berbeda dengan negara pada lintang rendah (tropis). Hal ini menyebabkan bahwa pemodelan untuk iklim dan prediksi hujan menjadi khas untuk suatu wilayah.

Berdasarkan penelitian yang berkaitan dengan pemodelan iklim dan prediksi curah hujan dengan pendekatan empiris, termasuk juga wilayah Indonesia,

masih menggunakan fitur cuaca/iklim secara parsial. Di sisi lain, secara teori bahwa iklim global, diantaranya *El Nino Southern Oscillation* (ENSO), *Index Oscillation Dipole* (IOD), *Indian Ocean Dipole* (IOD) dan *Dipole Mode Index* (DMI) secara teoritis sangat mempengaruhi curah hujan di Indonesia, utamanya wilayah Jawa dan Bali selain faktor lokal suatu wilayah [23], [28], [32], [33]. Kondisi atmosfer lokal dan faktor topografi suatu wilayah juga turut mempengaruhi curah hujan di wilayah tersebut [30], [34]–[36].

Kondisi topografi wilayah Indonesia yang bergunung, berlembah, serta banyak pantai, merupakan fenomena lokal dari negara lintang rendah sehingga model prediksi curah hujan menjadi cukup kompleks. Hal tersebut menyebabkan model iklim wilayah tropis yang mampu mensimulasikan dan memprediksi curah hujan di Indonesia secara efektif belum ditemukan [37].

Oleh karenanya dalam penelitian ini diusulkan suatu pendekatan baru model penduga curah hujan wilayah Zona Musim di Indonesia melalui kombinasi proses ekstraksi dan seleksi variabel prediktor serta penggabungan (*hybrid*) model *spatio-temporal* ARX dan ANN. Pendekatan dimaksud yaitu dengan melibatkan sejumlah fitur cuaca dan iklim dalam ruang lingkup lokal dan global dengan memperhitungkan faktor *time-lag* (*temporal*) dan lokasi (*spatial*).

Dengan pendekatan baru, baik dari segi prediktor yang simultan maupun model *hybrid*, prediksi curah hujan diharapkan hasilnya lebih akurat sehingga dapat digunakan sebagai informasi perencanaan di sejumlah bidang, antara lain pertanian, pertambangan, sumber daya air. Pemilihan prediktor dengan hanya melibatkan fitur relevan (signifikan) menjadikan proses komputasi lebih efisien sehingga informasi untuk pengambilan keputusan dapat dilakukan lebih cepat.

## 1.2 Perumusan Masalah

Permasalahan pada penelitian ini adalah:

- a. Bagaimana melakukan ekstraksi fitur tenggang waktu dari variabel prediktor yang berhubungan dengan variabel curah hujan?
- b. Bagaimana melakukan seleksi prediktor lokal dan global untuk model penduga curah hujan sehingga terpilih fitur yang relevan secara linier dan non linier?



- c. Bagaimana membangun model penduga curah hujan yang mampu mengintegrasikan variabel lokal/global, linier/non-linier, dan spatio-temporal?

### 1.3 Batasan Masalah

Dalam penelitian ini, data yang digunakan yaitu observasi cuaca harian yang diperoleh dari BMKG (<http://dataonline.bmkg.go.id/>) untuk wilayah Surabaya dan Malang, di Jawa Timur. Untuk wilayah Surabaya, meliputi Stasiun Meteorologi Kelas I, Juanda, Stasiun Meteorologi Perak I dan Stasiun Meteorologi Maritim Perak II. Sedangkan untuk wilayah Malang adalah Stasiun Geofisika kelas III Karangates dan Stasiun Klimatologi Kelas II Karang Ploso.

Data indeks iklim global periode harian dan per bulan diperoleh dari <http://climexp.knmi.nl>. Periode data observasi yang digunakan untuk mengembangkan model penduga curah hujan dan pengujian adalah periode Januari 1981 – Desember 2016 (36 tahun).

Variabel prediktor lokal yang digunakan dari data BMKG adalah minimum/mean/maksimum temperatur, kelembaban udara, lama penyinaran, kecepatan angin dan maksimum kecepatan angin. Sedangkan variabel prediktor global yang digunakan dari data KNMI adalah data anomali suhu permukaan laut ENSO (Nino1.2, Nino3, Nino3.4, Nino4) dan DMI (Dipole Mode Index). Model penduga dibangun dengan menggunakan perangkat lunak Matlab.

### 1.4 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah membangun model penduga curah hujan di wilayah Zona Musim melalui kombinasi proses ekstraksi dan seleksi variabel prediktor, penggabungan(*hybrid*) spatio-temporal ARX dan Jaringan Syaraf Tiruan untuk meningkatkan akurasi prediksi.

Proses ekstraksi dan seleksi fitur dimaksud adalah menemukan kandidat fitur *time-lags* dari prediktor lokal/global menggunakan metode DPCCA, dan melakukan seleksi prediktor baik hubungan linier maupun non linier menggunakan metode PCSUF. Hal ini bertujuan untuk memperoleh input prediktor yang relevan. Sedangkan model *hybrid* dibentuk melalui ST-ARX dan ANN dengan prediktor yang simultan dan bersifat *spatio-temporal*.

Manfaat dari penelitian ini adalah:

- a. Tersedianya metode seleksi prediktor yang mampu mengidentifikasi relevansi linier dan non linier terhadap variabel respon.
- b. Tersedianya model penduga curah hujan dengan untuk periode harian/bulanan yang melibatkan sejumlah variabel input secara simultan, baik variabel lokal/global, waktu maupun wilayah.
- c. Dapat digeneralisasi untuk wilayah ZoM di luar Jawa Timur untuk memprediksi curah hujan harian/per bulan, di mana 84% wilayah Indonesia adalah ZoM.

### **1.5 Kontribusi Penelitian**

Kontribusi penelitian ini adalah:

- a. Seleksi fitur yang memisahkan prediktor curah hujan berpola linier, non linier dan memilih sub himpunan prediktor yang relevan dan tidak redundan.
- b. Model penduga curah hujan hybrid ST-ARX dan ANN yang mampu mengintegrasikan input variabel lokal/global, linier/non-linier, dan spatio-temporal.

## **BAB 2**

### **KAJIAN PUSTAKA DAN DASAR TEORI**

#### **2.1 Kajian Pustaka**

##### **2.1.1 Penelitian terkait model penduga curah hujan**

Pemodelan curah hujan merupakan bagian penting dari penelitian bidang hidrologi dan lingkungan. Sejumlah penelitian terkait, antara lain peramalan aliran sungai dengan pendekatan ANN, ANFIS dan *Wavelet* [38], [39], [40], model penduga kekeringan dengan metode ANN dan ELM [41], [42], penguapan (*evaporation*) dengan metode *Relevance Virtual Machine* ELM, dan MARS [43], pengendapan (*precipitation*) dengan menggunakan metode ELM, ANFIS dan PSO-SVM [12], [44], [45], radiasi sinar matahari [46] dengan model markov, dan pemodelan temperatur udara dengan metode ANN [47], [48].

Model penduga curah hujan secara umum terbagi dua, yaitu model teoritis/konseptual dan model empiris/histori. Model empiris efektif untuk menemukan hubungan antara data input dan output ketika pengguna tidak memahami secara lengkap dari hubungan faktor-faktor fisik yang terjadi [3], cepat dalam pengembangan, dan kemudahan dalam implementasi waktu nyata [49]. Nilai akurasi prediksi curah hujan lebih diutamakan pada riset hidrologi. Di samping itu, pendekatan konseptual memerlukan proses komputasi yang mahal, kompleksitas tinggi dan padat peralatan *scientific* yang dalam beberapa tahun memerlukan pembaharuan peralatan [4], [49], [50].

Model penduga curah hujan berdasarkan data histori, antara lain dengan menggunakan ARIMA dengan berbagai variasinya, baik model autoregresi ARMA [18], ARIMA [32], maupun penggunaan variabel prediktor multivariat, yaitu VAR [19], ASTAR [32], SARIMAX [51], DARIMA dan NARX [15]. Metode penduga curah hujan berbasis ARIMA hanya sesuai untuk variabel dengan pola yang linier [52].

Pendekatan berbasis *machine learning* antara lain menggunakan ELM [48], [53], Fuzzy [54], GA [55], kNN [8], SVM [10], [56], [57], PSO [45], *Random Forest* [58] dan *Multiple Evaluation Criteria* [59]. Model ANN sebagai penduga

curah hujan, baik yang standar, pengembangan maupun penggabungan dengan metode lain (*hybrid*), cukup mendominasi. Model ANN standar, pengembangan pada tahap pra-proses dan cara representasi data, telah dilakukan untuk mendapatkan akurasi prediksi lebih baik [9], [47], [60], [61], [62], [63]. Demikian juga model *hybrid* ANN [3], [9], [13] [14], [38], [40], [51], [64], [65], [66], [67], [68], [69] menghasilkan akurasi yang cukup baik. Keunggulan lainnya dibandingkan ARIMA adalah bahwa ANN adaptif terhadap pola non-linier dan non-stasioner [3], [34], [53], [65], [65], [66].

Model peramalan curah hujan dengan studi kasus di Indonesia antara lain pada penelitian berikut:

- a. Pengembangan model peramalan curah hujan bulanan di wilayah Ngawi, Jawa Timur dengan metode *Adaptive Splines Threshold Autoregressive* (ASTAR), Fungsi Transfer dan model ARIMA dengan melibatkan prediktor data cuaca dan indeks global. Hasil terbaik adalah model Fungsi Transfer diikuti dengan ASTAR [32].
- b. Metode *Ensemble* berbasis ARIMA-ANFIS untuk peramalan curah hujan bulanan di wilayah Malang, dimana diperoleh hasil bahwa model individual ANFIS atau ARIMA memiliki kinerja lebih baik dibandingkan dengan *Ensemble* [72].
- c. Metode *Ensemble* berbasis *Wavelet*, ANFIS, ARIMA untuk peramalan total hujan bulanan di wilayah Indramayu, diperoleh hasil bahwa model *Ensemble* memiliki kinerja lebih baik dibandingkan dengan model individual pembentuknya, yaitu dengan nilai korelasi antara data riil dan prediksi adalah 0,72 [2].
- d. Metode gabungan *Fuzzy* dan Algoritma Genetika untuk peramalan curah hujan harian dengan prediktor data cuaca dengan studi kasus wilayah Jakarta. Diperoleh hasil bahwa model gabungan memberikan akurasi cukup baik untuk klasifikasi intensitas hujan harian, yaitu  $> 90\%$  [73]. Dengan studi kasus yang sama, apabila menggunakan metode *hybrid* GA dan *Partially Connected Feedforward Neural Network* (PCFNN), diperoleh nilai akurasi 81,52% [74].
- e. Metode VAR sebagai penduga curah hujan bulanan di sejumlah wilayah di Semarang (multi lokasi) dengan prediktor tunggal yang menghasilkan nilai

MAPE < 10%. Dibandingkan dengan ARIMA, akurasi VAR lebih baik [18], [19].

- f. Metode *Local Regression Smoothing* (LRS) dan *Fuzzy Grammatical Evolution* (GE) digunakan sebagai model penduga curah hujan bulanan di wilayah Bandung dengan nilai  $R^2=91,1\%$ . Apabila dibandingkan dengan ANFIS,  $R^2=80\%$  dan GE mempunyai  $R^2=74,35\%$  [75], [76].
- g. Metode *Exponential Smoothing Seasonal Planting Index* (ESSPI) untuk menduga curah hujan per bulan di wilayah Boyolali, dengan  $R^2=95,73\%$  [77].
- h. Metode *Seemingly Unrelated Regression* (SUR) untuk model GSTARIMA sebagai penduga curah hujan per bulan di wilayah Malang, dengan  $R^2=58\%$  [25].

### 2.1.2 Penelitian terkait pemilihan variabel prediktor

Model penduga curah hujan dengan variabel prediktor lokal (data cuaca) telah digunakan pada sejumlah penelitian. Variabel dimaksud antara lain suhu minimum/maksimum/rata-rata, kelembaban udara, lama penyinaran, arah dan kecepatan angin, dengan atau tanpa memperhitungkan pengaruh tenggang waktu dari variabel.

Variabel input dalam sejumlah penelitian prediksi curah hujan diantaranya adalah curah hujan dan aliran permukaan pada periode sebelumnya dengan seleksi variabel menggunakan regresi *stepwise* [3], curah hujan dan air tanah data bulanan pada periode sebelumnya yang ditentukan melalui *Partial MI* [13], curah hujan harian periode sebelumnya [14], curah hujan dan limpasan periode harian [61], [64], suhu minimum dan maksimum, curah hujan, limpasan, salju yang turun dengan periode data bulanan [63], curah hujan dan limpasan dengan periode harian dan bulanan [51], dan suhu dengan periode bulanan [66].

Sedangkan penggunaan variabel prediktor global (indeks iklim) sebagai input model penduga curah hujan, antara lain adalah *El-Nino Southern Oscillation* (ENSO) dan *Indian Ocean Dipole* (IOD) yang memperhitungkan pengaruh tenggang waktu berdasarkan nilai korelasi [65], *Southern Oscillation Index* (SOI), *Interdecadal Pacific Oscillation* (IPO), IOD, dan Niño 3.4 dengan hanya

memperhitungkan pengaruh tenggang waktu dari curah hujan satu bulan sebelumnya [78].

Penelitian curah hujan di suatu wilayah di Jawa Timur, menggunakan indeks ilkim bulanan, yaitu Nino3, Nino4, Nino3.4 dan DMI dengan melibatkan pengaruh tenggang waktu yang diidentifikasi dengan metode *prewhitening Cross-Correlation Function* (CCF) [32]. Penelitian serupa di wilayah Indramayu menggunakan *Sea-Surface Temperature* (SST) dan ENSO periode bulanan sebagai prediktor tanpa melibatkan pengaruh tenggang waktu [79].

Gabungan input lokal dan global sebagai prediktor dalam model penduga curah hujan telah digunakan dalam penelitian. Sebagai variabel input adalah suhu minimum, suhu maksimum, curah hujan periode bulan-bulan sebelumnya, SOI, IPO, DMI dan ENSO [60].

Pelibatan indeks iklim dan pengaruh tenggang waktu (*time-lags*) sebagai input dalam pemodelan curah hujan telah terbukti meningkatkan akurasi prediksi dalam sejumlah penelitian, diantaranya penggunaan ENSO, SST dan SOI untuk curah hujan di wilayah Australia [78], [80], termasuk juga pengaruh iklim global terhadap curah hujan di wilayah Indonesia [20], [29] [81], [82], [83].

### **2.1.3 Prediktor tenggang waktu, fitur relevan dan tidak redundan**

Humphrey dkk. [84] dalam studi bidang lingkungan menyatakan bahwa input penting dalam pemodelan yang harus dipertimbangkan adalah ruang dan waktu, hubungan kausalitas variabel observasi pada lokasi yang berbeda, dan pengaruh tenggang waktu dari variabel yang sama. Akibatnya, jumlah input yang potensial menjadi sangat banyak. Tantangannya bahwa seleksi variabel adalah memilih jumlah variabel input seminimal mungkin yang mewakili karakteristik hubungan input-output sekaligus meminimalkan redundansi variabel.

Model penduga curah hujan sebagai bagian dari studi lingkungan merupakan salah satu aplikasi dari peramalan data runtun waktu (*time series*) sehingga keterlibatan dari variabel tenggang waktu akan mempengaruhi akurasi prediksi. Sejumlah cara untuk mengidentifikasi pengaruh dan hubungan tenggang waktu terhadap variabel respon telah banyak digunakan, antara lain teknik korelasi-silang [44], auto-korelasi [54], [55], [85], [86] [87], [88]. Keduanya hanya dapat

mengidentifikasi hubungan variabel tenggang waktu yang pengaruhnya bersifat linier dan stasioner.

Dalam penelitian lain, kandidat prediktor tenggang waktu untuk memprediksi curah hujan tidak ditentukan melalui metode tertentu, tetapi dengan menetapkan *lags*  $t-1$  sampai dengan  $t-k$  untuk mendapatkan kombinasi input terbaik [53], [60], [64], [66]. Cara lainnya adalah memilih fitur yang signifikan dengan metode tertentu, antara lain *Stepwise Regression* [3], [89] dan MI/PMI [14], [48], [57], [90]. Keunggulan dari MI adalah kemampuannya mendeteksi variabel input dengan pengaruh non-linier [5].

Metode lain dengan kemampuan lebih baik dibanding auto-korelasi dan korelasi silang yang mampu mengidentifikasi pengaruh tenggang waktu bersifat linier, non linier dan non stasioner adalah *Detrended Partial Cross-Correlation Analysis* (DPCCA) [91], [92], [93].

Untuk memperoleh efisiensi dalam model penduga data runtun waktu, tidak semua variabel prediktor yang telah diidentifikasi, akan digunakan sebagai input dalam model. Diperlukan proses pemilihan variabel input yang berguna secara langsung (*relevant*), namun tidak berlebihan (*redundant*). Sejumlah metode untuk seleksi variabel input dari model penduga deret berkala telah diperkenalkan, diantaranya adalah berbasis metode korelasi [78]–[81] untuk memilih variabel input dengan pola keterkaitan yang linier, *Symmetrical Uncertainty* (SU) [98], [99], [100] untuk pola yang non-linier, MI [48], [101], [102], [103] untuk pola yang non-linier, *Principle Component Analysis* (PCA) [104], PSO [105], *Entropy* [8], ANFIS [12], maupun *Wavelet* dengan pendekatan model *hybrid* [49].

#### **2.1.4 Penelitian spatio-temporal dan model *hybrid***

Sejumlah penelitian baru telah memperkenalkan pendekatan yang berbeda untuk model penduga data runtun waktu multivariat, yaitu menggunakan *spatio-temporal*. Faghmous dan Kumar [36] menyarankan bahwa untuk meningkatkan akurasi, skalabilitas dan kemudahan interpretasi dari persoalan deret berkala multivariat, dapat digunakan *Spatio-Temporal Data Mining* (STDM). Data curah hujan yang bersifat *zero-inflated* akan lebih fleksibel dalam analisisnya jika

dimodelkan sebagai persoalan STDM [34]. Dengan menggunakan data curah hujan mingguan di wilayah Brasil, pendekatan STDM menghasilkan MSE lebih baik.

Penelitian spatio-temporal di bidang hidrologi antara lain dilakukan pada peramalan vegetasi untuk mengevaluasi dampak perubahan iklim. Dengan menggunakan multiplikatif Spatio-Temporal ARIMA, model ini menghasilkan kinerja lebih baik dibanding ARIMA [106]. Studi kasus pemodelan curah hujan harian dengan pendekatan spatio-temporal telah dilakukan di wilayah Jawa Timur untuk menganalisis hujan ektrim [30]. Pendekatan spatio-temporal dapat juga digunakan pada model konseptual, yaitu model *Empirical Orthogonal Function* untuk analisis hujan harian di sejumlah wilayah di Jawa Timur [107]. Pendekatan spatio-temporal juga digunakan untuk model peramalan dengan metode ANN [108], [109].

Pengembangan model penduga untuk data runtun waktu, saat ini menuju kepada penggabungan dua atau lebih metode. Model *hybrid* untuk prediksi curah hujan diantaranya adalah *Autoregressive-ANN* [38], *wavelet-ANN* [46], *wavelet-ANFIS* [16], [59], *DARIMA-NARX* [15], *Soft Computing-MARS* [110], *wavelet-Gene Expression Programming* [55], *PSO-SVM* [45], *SOM-NARX* [13], *CAPSO-ANN* [111], *GA-LM-ANN* [3], dan *Multiple Regression-ANN* [65].

Secara umum, pengembangan model peramalan data runtun waktu berdasarkan penelitian Makridakis dan Hibon [112] antara lain disimpulkan bahwa:

- a. Metode yang canggih atau rumit secara statistik tidak menjamin hasil peramalan yang lebih akurat daripada metode yang sederhana. Namun hal ini tidak berarti bahwa beberapa metode yang canggih tidak menghasilkan kinerja yang baik karena tidak ada batas yang jelas atas klasifikasi metode disebut sederhana atau canggih.
- b. Keakuratan hasil atas kombinasi sejumlah metode lebih baik dibanding metode tunggal (individual) pada proses peramalan, baik satu periode maupun sejumlah periode ke depan.



## 2.2 Dasar Teori

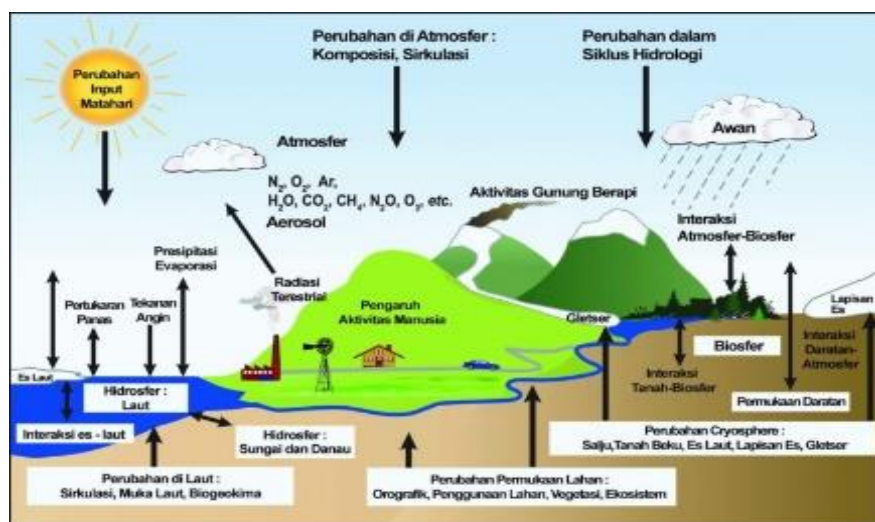
### 2.2.1 Cuaca dan iklim

Cuaca adalah keadaan/fenomena fisik dari atmosfer yang berhubungan dengan suhu, tekanan udara, angin, awan, kelembaban udara, radiasi, dan jarak pandang di suatu tempat dan pada waktu tertentu (per jam atau harian). *Iklim* adalah aspek dari cuaca di suatu tempat dan pada waktu tertentu dalam jangka waktu panjang. Misalnya, evaluasi dan prakiraan hujan bulanan, prakiraan musim hujan dan kemarau. Sedangkan musim adalah pembagian iklim dalam rentang waktu satu tahun [113].

Iklim terdiri atas atmosfer daratan, salju dan es, lautan dan badan-badan air lainnya, serta makhluk hidup. Iklim merupakan suatu sistem yang kompleks dan interaktif. Sistem berevolusi terhadap waktu oleh adanya pengaruh-pengaruh yang berasal dari dinamika internalnya dan juga sebagai akibat terjadinya perubahan pada faktor-faktor luar yang mempengaruhi iklim. Iklim umumnya didefinisikan sebagai cuaca rata-rata, dan dengan definisi yang demikian perubahan iklim dan cuaca saling terkait. Meskipun cuaca dan iklim terkait erat, terdapat perbedaan-perbedaan penting. Cuaca memiliki perilaku *chaotic* yang membuat tidak dapat diprediksi setelah beberapa hari ke depan. Proyeksi perubahan iklim (yakni cuaca rata-rata jangka panjang) akibat perubahan komposisi atmosfer atau akibat faktor-faktor lain adalah hal yang sangat berbeda dan sangat mungkin untuk dilakukan [114]. Sistem iklim dijelaskan pada Gambar 2.1.

Berikut dijelaskan definisi dan pengertian sejumlah variabel terkait dengan meteorologi dan iklim [115]. Curah hujan merupakan ketebalan air hujan yang terkumpul pada luasan  $1 \text{ m}^2$ . Hujan adalah titik-titik air di udara atau awan yang sudah terlalu berat karena kandungan airnya sudah sangat banyak, sehingga akan jatuh kembali ke permukaan bumi sebagai hujan (presipitasi). Satuan curah hujan yang digunakan adalah mm.

Suhu udara adalah ukuran energi kinetik rata-rata dari pergerakan molekul-molekul. Dalam sistem dua benda, benda yang kehilangan panas dikatakan benda yang bersuhu lebih tinggi. Lama penyinaran matahari merupakan salah satu dari unsur iklim, dan didefinisikan sebagai kekuatan matahari yang melebihi  $120 \text{ w/m}^2$ .



Gambar 2.1 Unsur-unsur cuaca dan iklim [139]

Kelembaban udara adalah jumlah kandungan uap air yang ada dalam udara. Kandungan uap air di udara berubah-ubah bergantung pada suhu. Makin tinggi suhu, makin banyak kandungan uap airnya. Kelembapan relatif (nisbi), yaitu angka dalam persen yang menunjukkan perbandingan antara banyaknya uap air yang benar-benar dikandung udara pada suhu tertentu dan jumlah uap air maksimum yang dapat dikandung udara.

Secara umum wilayah Indonesia dikenal sebagai wilayah tropis yang memiliki 2 musim, yaitu kemarau dan hujan. Namun berdasarkan sistem perubahan angin musiman (monsun) berkaitan erat dengan sistem penghujan, musim dibagi menjadi 4 yaitu musim hujan (Desember-Februari), musim peralihan hujan-kering/transisi I (Maret-Mei), musim kering (Juni-Agustus), dan musim peralihan kering-hujan/transisi II (September-November). Sedangkan untuk tipe iklim, seringkali dikelompokkan menjadi sejumlah tipe, yaitu sangat basah, lembab, agak kering, dan sangat kering [116].

Untuk keperluan prakiraan musim, BMKG membagi wilayah Indonesia menjadi menjadi 2 (dua) kelompok yaitu daerah zona musim (ZoM) dan daerah non zona musim (non ZoM). ZoM didefinisikan sebagai daerah yang pola hujan rata-ratanya memiliki perbedaan yang jelas antara periode musim kemarau dan musim hujan. Daerah-daerah yang pola hujan rata-ratanya tidak memiliki perbedaan yang jelas antara periode musim kemarau dan musim hujan, disebut Non ZoM. Secara

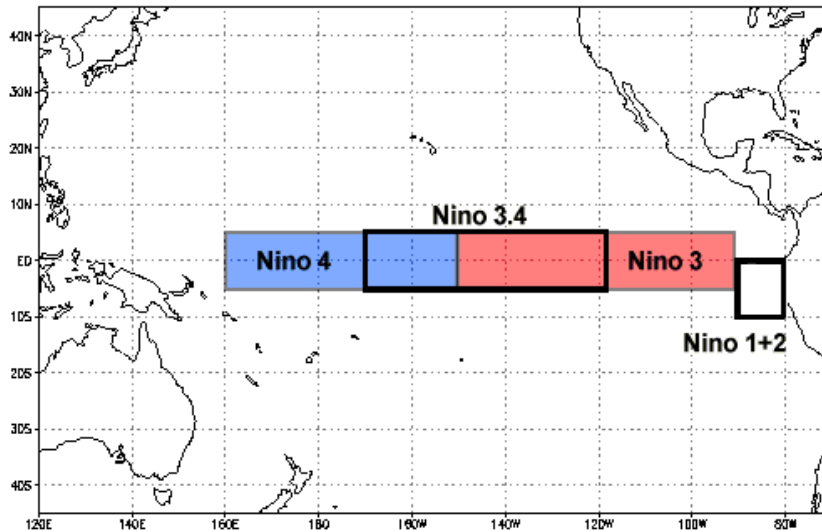
klimatologis, wilayah Indonesia terdapat 407 pola hujan dimana 342 diantaranya adalah ZoM. Wilayah Jawa Timur merupakan daerah yang memiliki pola monsun dan dibagi dalam 60 daerah ZoM, dimana tiap zona dibatasi oleh garis berwarna biru dan berlabel nomer zona, sesuai pada Gambar 2.2 [117].



Gambar 2.2 Zona Musim di Jawa Timur (sumber: web BMKG)

Hujan di Indonesia dipengaruhi oleh iklim global, diantaranya ENSO (*El Nino-Southern Oscillation*) di mana besarnya pengaruh beragam. Pengaruh itu sangat besar pada daerah yang memiliki pola hujan monsoon. El Nino merupakan salah satu bentuk penyimpangan iklim di Samudra pasifik yang ditandai dengan kenaikan SST (*Sea Surface Temperature*) di daerah katulistiwa bagian tengah dan timur. Sebagai indikator untuk memantau kejadian *El Nino* dan *La Nina*, digunakan data pengukuran SST di zona lautan pasifik, seperti pada Gambar 2.3 [27], [118].

Bentuk penyimpangan iklim yang menyerupai *El Nino* dan *La Nina* adalah IOD (*Indian Ocean Dipole*), merupakan anomali SST India yang berkontribusi pada kekeringan di Indonesia. IOD diukur dengan suatu indeks yang dinamakan *Dipole Mode Index* (DMI), yaitu perbedaan SST di bagian barat samudra hindia ( $50^{\circ} - 70^{\circ}$  BT,  $10^{\circ}$  LS –  $10^{\circ}$  LU) dan di bagian timur samudra hindia ( $90^{\circ} - 110^{\circ}$ ,  $10^{\circ}$  LS – ekuator).



Gambar 2.3 Zona ENSO di samudra pasifik [118]

Anomali dari SST diukur berdasarkan indeks ENSO. Nino1\_2 adalah wilayah pengamatan anomali SST pada koordinat  $0^{\circ}$ – $10^{\circ}$  LS dan  $90^{\circ}$ – $80^{\circ}$  BB, Nino3 pada koordinat  $5^{\circ}$  LU– $5^{\circ}$  LS dan  $150^{\circ}$ – $90^{\circ}$  BB, Nino3.4 pada  $5^{\circ}$  LU– $5^{\circ}$  LS,  $160^{\circ}$ BT– $150^{\circ}$ BB, serta Nino4 pada  $5^{\circ}$  LU– $5^{\circ}$  LS dan  $170^{\circ}$  BB– $120^{\circ}$  BB.

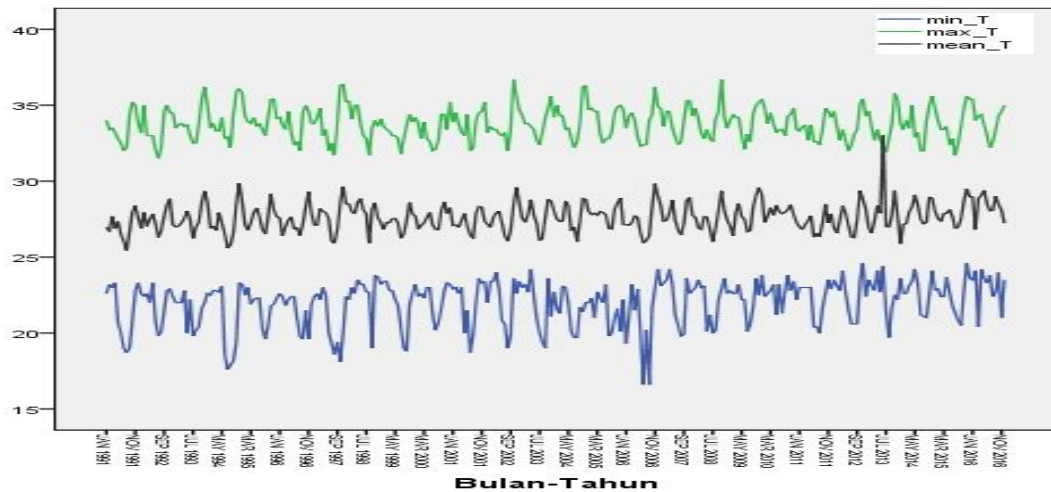
### 2.2.2 Data runtun waktu multivariat

Runtun waktu adalah representasi data yang terurut berdasarkan waktu kejadian dengan interval waktu yang sama di mana antar nilai saling berhubungan [119]. Data runtun waktu dapat dinotasikan sebagai  $X_i(t)$ ;  $[i = 1, 2, 3, \dots, n, t = 1, 2, 3, \dots, m]$ . Jika nilai  $n \geq 2$ , disebut sebagai runtun waktu multivariat, sementara apabila  $n = 1$  dinamakan *univariate time series* (UTS).

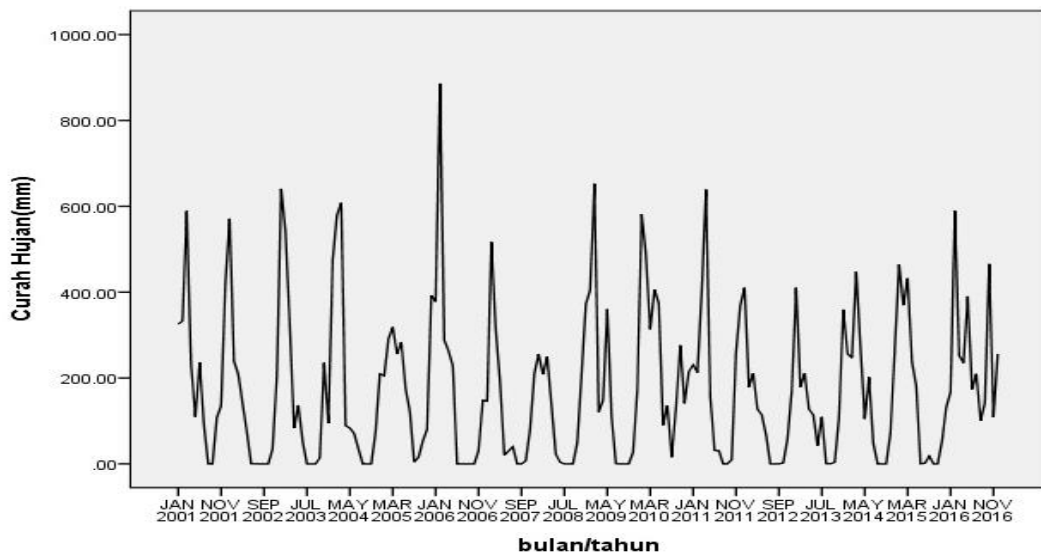
Tipe data runtun waktu dibedakan atas 2 jenis, yaitu stasioner dan non-stasioner. Yang dimaksud dengan UTS stasioner adalah bahwa variasi dan fluktuasi nilainya berada pada sekitar nilai rata-rata (relatif konstan). Data yang disajikan pada Gambar 2.4 yaitu suhu minimum( $^{\circ}$ C), suhu rata-rata( $^{\circ}$ C) dan suhu maksimum( $^{\circ}$ C) adalah data runtun waktu multivariat yang termasuk pada tipe stasioner. Data bertipe stasioner dapat memiliki pola musiman jika terjadi perulangan pola yang relatif sama pada periode berikutnya. Contoh pada Gambar 2.5, yaitu data curah hujan periode bulanan (mm/bulan) merupakan data stasioner

yang memiliki pola musiman. Sedangkan data indeks iklim anomali SST pada Gambar 2.6 adalah bertipe non-stasioner.

Yang dimaksud dengan runtun waktu non-linier adalah data hasil proses stokastik yang tidak dapat dimodelkan dengan pendekatan linier. Sifat non-linier terjadi pada data runtun waktu stasioner maupun non-stasioner.



Gambar 2.4 Contoh data runtun waktu meteorologi [120]



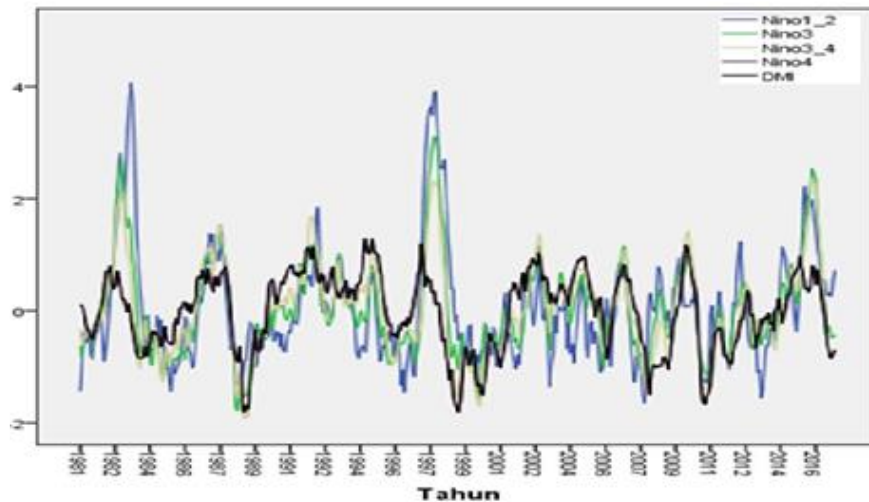
Gambar 2.5 Plot data runtun waktu curah hujan [120]

Salah satu alat analisis untuk mengidentifikasi data time-series bersifat stasioner, musiman dan adanya pengaruh dari tenggang waktu (*time-lags*) adalah fungsi ACF (*Autocorrelation Function*) dan PACF (*Partial ACF*). Gambar 2.7 adalah *correlogram* dari ACF dan PACF data curah hujan bulanan.

Prosedur pengujian suatu data runtun waktu adalah bersifat stasioner, metode *augmented dickey-fuller test* (*adf-test*) dapat digunakan sebagai salah satu alat uji [121]. Persamaan untuk *adf-test* adalah sebagai berikut:

$$\Delta Y_t = \beta_1 + \beta_2 t + \gamma Y_{t-1} + \alpha \sum_{t=1}^n \Delta Y_t + \varepsilon_t \quad (1)$$

$$\text{dimana } e_t = \Delta Y_{t-1} = (Y_{t-1} - Y_{t-2}), \Delta Y_{t-2} = (Y_{t-2} - Y_{t-3}) \quad (2)$$

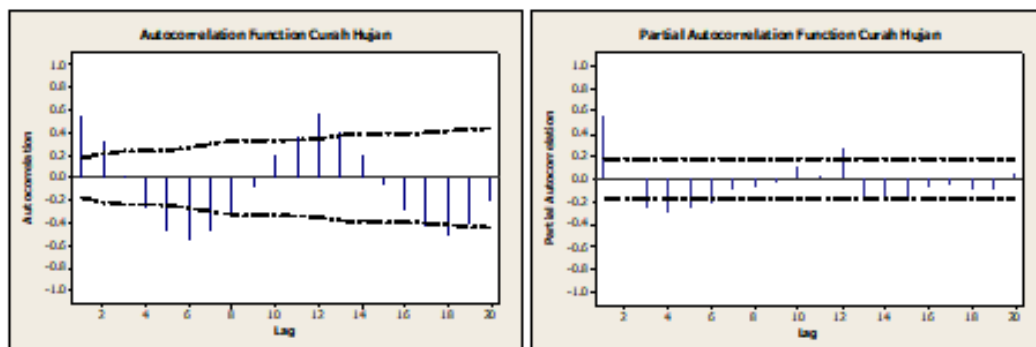


Gambar 2.6 Data runtun waktu non-stasioner [120]

Dengan hipotesis nol bahwa data runtun waktu adalah stasioner, nilai uji *adf* adalah  $\hat{\gamma}/(\text{SE}(\hat{\gamma}))$ . Semakin negatif nilai uji *adf*, maka data runtun waktu adalah stasioner. Sedangkan prosedur untuk menguji non-linieritas, digunakan metode *white neural net test* [122]. Apabila terdapat hubungan rekursif berikut:

$$t(n+1) = f(t(n)) + e \quad (3)$$

dimana  $e$  adalah kesalahan acak,  $t(n)$  dan  $t(n+1)$  adalah data runtun waktu pada saat  $n$  dan  $(n+1)$ , dan  $f$  adalah suatu fungsi, maka *white-test* akan menguji bahwa  $f$  adalah



Gambar 2.7 ACF dan PACF dari curah hujan di Surabaya [120]

fungsi linier  $f(x) = ax + b$ . Hipotesis data runtun waktu adalah linier akan diterima jika nilai uji *white* kurang dari nilai kritis pada tabel *Chi-Square*.

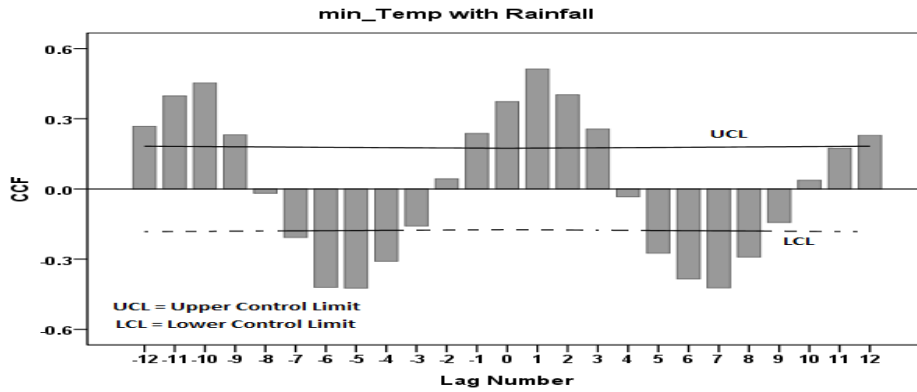
### 2.2.3 Ekstraksi fitur tenggang waktu (*time-lags*)

Untuk mengidentifikasi hubungan linier dua variabel berdasarkan tenggang waktu (*time-lags*) tertentu, alat analisis yang umum digunakan adalah *Cross Correlation Function* (CCF). Untuk menghitung CCF digunakan persamaan 4.

$$r_{xy}(T) = \frac{\sigma_{xy}(T)}{\sqrt{\sigma_{xx}(0)}\sqrt{\sigma_{yy}(0)}} \quad (4)$$

di mana  $r_{xy}(T)$  adalah korelasi variabel  $x$  dan variabel  $y$  pada periode ke  $T$  dan  $\sigma_{xy}$  adalah kovariansi( $x, y$ ),  $\sigma_{xx}$  adalah variansi( $x$ ) dan  $\sigma_{yy}$  adalah variansi( $y$ ).

Dalam hubungan antara dua runtun waktu ( $y_t$  dan  $x_t$ ), data  $y_t$  mungkin berhubungan dengan *lags* yang lalu dari data  $x$ . Sampel CCF membantu mengidentifikasi *lags* dari variabel  $x$  yang mungkin merupakan kandidat prediktor dari  $y_t$ . Sampel CCF didefinisikan sebagai himpunan korelasi sampel antara  $x_{t+h}$  dan  $y_t$  untuk  $h = 0, \pm 1, \pm 2, \pm 3$ , dan seterusnya.



Gambar 2.8 Correlogram CCF antara curah hujan dan suhu minimum [133]

Nilai negatif dari  $h$  menunjukkan bahwa korelasi antara variabel  $x$  pada waktu sebelum  $t$  dan variabel  $y$  pada saat  $t$ . Hasil nilai korelasi dapat diuji dengan membandingkan nilai korelasi silang dengan *standard error*  $\pm 1/\sqrt{n-k}$ . Pada Gambar 2.8, ditunjukkan *correlogram* CCF curah hujan dengan suhu minimum. Apabila nilai dari korelasi lebih dari batas UCL dan LCL, maka korelasinya signifikan.



Karena CCF memiliki keterbatasan hanya untuk identifikasi pengaruh tenggang waktu yang bersifat linier dan stasioner, cenderung redundan dan bersifat *greedy* [123], maka metode lain dikembangkan, yaitu DCCA/DPCCA. Keunggulan metode ini adalah kemampuan identifikasi hubungan diantara dua variabel yang bersifat linier/non-linier dan non-stasioner sehingga dapat digunakan untuk memperoleh kandidat variabel prediktor lebih lengkap.

Langkah-langkah DPCCA adalah sebagai berikut [91], [92]:

1. Jika terdapat  $m$  variabel runtun waktu,  $i = 1, 2, \dots, N$ ;  $\{x_i^1\}, \{x_i^2\}, \dots, \{x_i^m\}$
2. Masing masing *runtun waktu* dianggap sebagai *random-walk* :

$$P_k^j \equiv \sum_{i=1}^k x_i^j \text{ untuk } j = 1, 2, \dots, m; k = 1, 2, \dots, N \quad (5)$$

3. Buat profil  $P$  sebanyak  $(N-s)$ . Setiap kotak berisi nilai  $(s+1)$  dari  $i$  s/d  $(s+i)$ .
4. Hitung *trend* lokal dengan *polynomial fit*.
5. Hilangkan *trend* dari *random walk*

$$Y_{(i-1)(s+1)+k-i+1}^j = P_k^i - \widetilde{P}_{k,i}^j; \quad \widetilde{P}_{k,i}^j (i \leq k \leq i+s) \quad (6)$$

Diperoleh nilai residual dari runtun waktu

$$Y_l^j, l = 1, 2, \dots, (N-s)(s+1) \text{ untuk setiap series } \{x_i^j\}$$

6. Hitung Kovariansi semua pasangan residual

$$F_{j_1, j_2}^2(s) \equiv \frac{\sum_{l=1}^{(N-s)(s+1)} Y_l^{j_1} Y_l^{j_2}}{(N-s)(s-1)} \quad (7)$$

Diperoleh nilai konvariansi

$$F^2(s) = \begin{pmatrix} F_{1,1}^2(s) & F_{1,2}^2(s) & \dots & F_{1,m}^2(s) \\ F_{2,1}^2(s) & F_{2,2}^2(s) & \dots & F_{2,m}^2(s) \\ \vdots & \vdots & \ddots & \vdots \\ F_{m,1}^2(s) & F_{m,2}^2(s) & \dots & F_{m,m}^2(s) \end{pmatrix} \quad (8)$$

7. Hitung nilai korelas silang untuk pasangan data runtun waktu

$$\rho_{j_1, j_2}(s) \equiv \frac{F_{j_1, j_2}^2(s)}{F_{j_1, j_1}(s) F_{j_2, j_2}(s)} \quad (9)$$

$$\rho(s) = \begin{pmatrix} \rho_{1,1}(s) & \rho_{1,2}(s) & \dots & \rho_{1,m}(s) \\ \rho_{2,1}(s) & \rho_{2,2}(s) & \dots & \rho_{2,m}(s) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m,1}(s) & \rho_{m,2}(s) & \dots & \rho_{m,m}(s) \end{pmatrix} \quad (10)$$



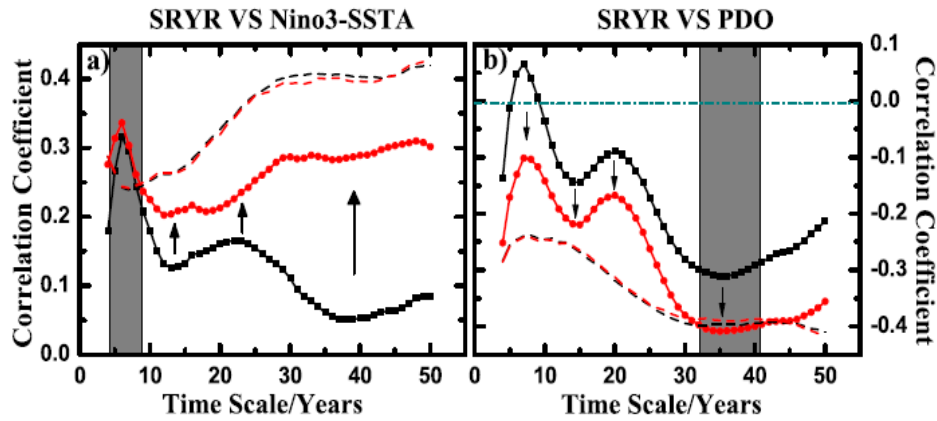
8. Hitung matriks invers dari korelasi silang

$$C(s) = \rho^{-1}(s) = \begin{pmatrix} C_{1,1}(s) & C_{1,2}(s) & \dots & C_{1,m}(s) \\ C_{2,1}(s) & C_{2,2}(s) & \dots & C_{2,m}(s) \\ \vdots & \vdots & \vdots & \vdots \\ C_{m,1}(s) & C_{m,2}(s) & \dots & C_{m,m}(s) \end{pmatrix} \quad (11)$$

9. Hitung DPCCA

$$\rho_{DPCCA}(j_1, j_2, s) \equiv \frac{-C_{j_1, j_2}(s)}{\sqrt{C_{j_1, j_1}(s)C_{j_2, j_2}(s)}} \quad (12)$$

Gambar 2.9 adalah contoh hasil DPCCA (garis berwarna merah) bila dibandingkan dengan DCCA (garis berwarna hitam) untuk hubungan antara curah hujan dengan Nino3 dan curah hujan dengan *Pasific Decadal Oscilation* (PDO).



Gambar 2.9 DPCCA antara curah hujan, Nino3 dan PDO [91]

#### 2.2.4 Seleksi fitur

Dari studi Yu dkk. [98], fitur prediktor untuk pemodelan MTS dibagi menjadi 3 subset yang saling lepas, yaitu relevan yang sangat kuat, relevan yang lemah dan fitur yang tidak relevan. Subset pertama mutlak/harus ada sebagai prediktor dari variabel respon dan tidak dapat dihapus karena akan berpengaruh kepada hasil prediksi. Subset kedua kurang diperlukan, kecuali pada kondisi

tertentu. Subset ketiga tidak diperlukan sebagai prediktor pada kondisi apapun. Pembagian fitur prediktor atas 3 subset dijelaskan pada Gambar 2.10.

Untuk melakukan seleksi prediktor yang relevan, asosiasi antar fitur diasumsikan linier. Suatu prediktor disebut relevan terhadap variabel respon jika korelasinya signifikan. Formula untuk menentukan prediktor relevan adalah dengan menggunakan korelasi Pearson, yang ditulis pada persamaan 13.

$$\rho_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (13)$$

$\rho_{xy}$  adalah korelasi antara variabel x dan variabel y dan nilainya signifikan jika lebih dari nilai kritis tabel korelasi dengan derajat bebas dan level signifikansi tertentu.

Sedangkan untuk mengevaluasi asosiasi (hubungan) antar fitur yang bersifat non linier, digunakan *symmetrical uncertainty* (SU). Formula SU berbasis pada *entropy* dan *information gain* [94]. Jika X adalah suatu variabel, maka *entropy* didefinisikan sesuai persamaan 14.

$$H(X) = -\sum_i P(x_i) \log(P(x_i)) \quad (14)$$

dan *entropy* X setelah ada nilai observasi dari variabel lain, didefinisikan sebagai persamaan 15.

$$H(X|Y) = -\sum_j P(y_j) \sum_i P(x_i|y_j) \log(P(x_i|y_j)) \quad (15)$$

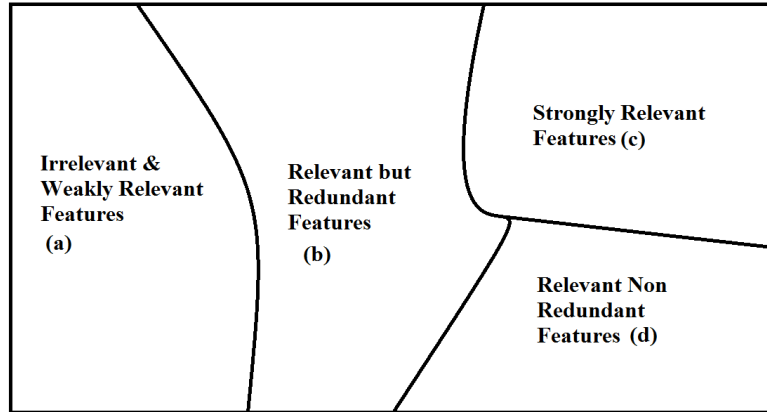
di mana  $P(x_i)$  adalah probabilitas *prior* untuk semua nilai dari X, dan  $P(x_i / y_j)$  adalah probabilitas *posterior* dari X ketika diberikan nilai dari Y. Jumlahan *entropy* X mengikuti persamaan 16.

$$IG(X|Y) = H(X) - H(X|Y) \quad (16)$$

Namun, *IG* bersifat bias sehingga perlu dilakukan normalisasi sesuai persamaan 17 sebagai *Symmetrical Uncertainty* (SU).

$$SU(X, Y) = 2[IG(X|Y) / (H(X) + H(Y))] \quad (17)$$

$SU(X, Y)$  antara variabel X dan Y disebut signifikan jika nilainya lebih dari nilai kritis tabel *Chi-Square* dengan derajat bebas *significant level*  $\alpha$  [124].



Gambar 2.10 Fitur prediktor yang relevan dan tidak relevan [98]

### 2.2.5 Model ARMA, ARIMA dan ARIMAX

Salah satu model penduga data runtun waktu yang populer dan paling sering digunakan adalah model ARIMA. Model AR menunjukkan bahwa terdapat relasi antara nilai sekarang ( $Z_t$ ) dan nilai pada periode sebelumnya ( $Z_{t-k}$ ) dengan tambahan nilai random  $e_t$  [125]. Model MA menunjukkan bahwa ada relasi antara nilai sekarang ( $Z_t$ ) dengan residual pada periode lalu ( $a_{t-k}$ , dengan  $k=1, 2, \dots$ ). Model ARIMA( $p, d, q$ ) adalah campuran dari AR( $p$ ) dan MA( $q$ ) dengan pola non-stasioner dan  $d$  adalah level *differencing*. Bentuk ARIMA( $p, d, q$ ) adalah sesuai dengan persamaan 18.

$$\phi_p(1 - B)^d Z_t = \theta_q(B) a_t \quad (18)$$

di mana  $p$  adalah order dari model AR dan  $q$  adalah order dari model MA dengan level *difference*  $d$

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 \dots - \phi_p B^p) \quad (19)$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q) \quad (20)$$

Bentuk umum dari model ARIMA dengan pola data musiman ditulis dengan ARIMA( $p \ d \ q$ ) ( $P \ D \ Q$ )<sup>s</sup> mengikuti persamaan (21).

$$\phi_p(B) \Phi_P(B^s) (1 - B)^d (1 - B^s)^D Z_t = \theta_q(B) \Theta_Q(B^s) a_t \quad (21)$$

di mana  $s$  adalah periode musiman. Persamaannya adalah sebagai berikut:

$$\Phi_P(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} \dots - \Phi_P B^{Ps}) \quad (22)$$

$$\text{dan } \Theta_Q(B) = (1 - \Theta_1 B^s - \Theta_2 B^{2s} \dots - \Theta_Q B^{Qs}). \quad (23)$$

Flow-chart untuk pemodelan ARIMA dijelaskan pada Gambar 2.11.

ARIMAX adalah pengembangan dari ARIMA dengan ditambahkan variabel eksternal (*eXogenous Variable*) sebagai penduga terhadap variabel respon [32], [126]. Salah satu bentuk ARIMAX fungsi transfer satu input adalah persamaan 24.

$$Z_t = C + \frac{\omega_s(B)}{\delta_r(B)} B^b X_t + N_t \quad (24)$$

di mana:

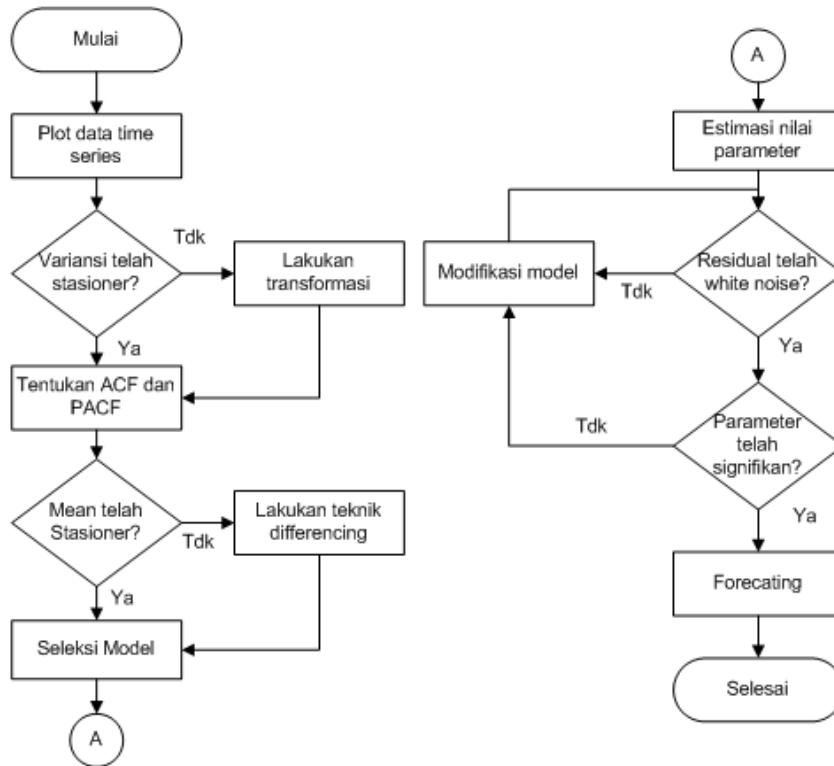
$$N_t = \frac{\theta(B)}{\phi(B)} a_t \quad (25)$$

$$\omega_s(B) = \omega_0 - \omega_1 B - \omega_2 B^2 \dots - \omega_s B^s \quad (26)$$

$$\delta_r(B) = \delta_0 - \delta_1 B - \delta_2 B^2 \dots - \delta_r B^r \quad (27)$$

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 \dots - \phi_p B^p) \quad (28)$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q) \quad (29)$$



Gambar 2.11 Flow-chart pemodelan ARIMA [120]

### 2.2.6 Model artificial neural networks (ANN)

ANN adalah sistem pemrosesan informasi yang mempunyai karakteristik kinerja tertentu dengan mengadaptasi dari jaringan syaraf biologi. ANN telah digeneralisasi sebagai model matematik dari kognisi manusia, dengan didasarkan pada asumsi-asumsi sebagai berikut:

1. Pemrosesan informasi terjadi pada banyak elemen-elemen sederhana yang disebut dengan *Neurons*.
2. Sinyal-sinyal dikirim antar neuron melalui *connection-links*.
3. Setiap *connection-links* mempunyai bobot yang sesuai, tergantung tipe jaringan Neural.
4. Setiap Neuron menggunakan *activation function* (biasanya non linier) yang merupakan jumlahan dari sinyal-sinyal input untuk menentukan sinyal output.

ANN dibedakan oleh:

1. Pola koneksi antar neuron (disebut dengan *architecture*).
2. Metode penentuan bobot pada koneksi-koneksi (disebut dengan *training, learning, atau algorithm*).
3. *Activation function*.

ANN terdiri atas banyak elemen yang disebut neuron, unit, sel atau node saling terhubung satu dengan yang lain melalui *synaptic links* yang mempunyai bobot yang berkaitan. Bobot mewakili informasi yang akan digunakan oleh jaringan untuk menyelesaikan persoalan. ANN dapat diterapkan pada berbagai tipe persoalan, misalnya untuk klasifikasi pola, membuat pemetaan dari pola input, prediksi nilai mendatang berdasarkan data masa lalu, ataupun mencari solusi dari permasalahan optimasi dengan banyak konstrain.

Apabila model ARIMA mempunyai keterbatasan hanya mampu memprediksi data runtun waktu dengan pola dan fluktuasi linier, sebaliknya ANN memiliki kemampuan untuk menggambarkan berbagai pola non linier dari data. Satu kelebihan yang signifikan dari model ANN sebagai pendekatan pola non linier dibandingkan metode yang lain adalah bahwa ANN merupakan pendekatan umum untuk fungsi kelas yang besar dengan akurasi yang tinggi [127].

Arsitektur lapisan tersembunyi tunggal *feedforward* paling banyak digunakan untuk pemodelan dan prediksi data runtun waktu. Model ini memiliki karakteristik

sederhana dengan tiga lapis pemroses pada unit-unit yang terhubung dengan *acyclic links*. Hubungan antara output ( $y_t$ ) dan input ( $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ ) dinyatakan dalam persamaan 30.

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i}) + \varepsilon_t \quad (30)$$

di mana  $\alpha_j$ , ( $j=0, 1, 2, \dots, q$ ) dan  $\beta_{ij}$  ( $i=0, 1, 2, \dots, p$ ), ( $j=1, 2, \dots, q$ ) adalah parameter yang disebut bobot koneksi,  $p$  adalah jumlah *node input* dan  $q$  adalah jumlah *node tersembunyi*. Fungsi logistik seringkali digunakan pada lapisan tersembunyi, dengan persamaan 31.

$$g(x) = \frac{1}{1+e^{-x}} \quad (31)$$

Model ANN dari persamaan (31) membentuk fungsi pemetaan non linier dari pengamatan lampau ( $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ ) kepada nilai mendatang  $y_t$  dinyatakan sebagai persamaan (32).

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w) + \varepsilon_t \quad (32)$$

di mana  $w$  adalah vektor untuk semua parameter dan  $f$  adalah fungsi yang ditentukan oleh struktur *network* dan bobot koneksi. Persamaan (32) berimplikasi satu *node output* pada lapisan output

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}, w) + \varepsilon_t \quad (33)$$

### 2.2.7 Model *hybrid* ARIMA dan ANN

Antara kedua model ARIMA dan ANN telah berhasil dalam domain yang berbeda, linier atau non linier. Namun tidak ada dari keduanya merupakan model yang sesuai untuk semua kondisi. Pendekatan ARIMA untuk memodelkan permasalahan non linier yang kompleks tidak cukup baik. Sebaliknya, pemodelan ANN untuk permasalahan linier menghasilkan kinerja yang beragam, bergantung pada karakteristik data, antara lain adanya multikolinieritas, pencilan, jumlah sampel dan tingkat *noise* [127].

Karenanya tidak disarankan mengaplikasikan ANN pada beragam tipe data. Hal ini karena dalam persoalan nyata, akan sulit mengenali secara lengkap karakteristik data sehingga metodologi *hybrid* yang menggabungkan kemampuan pemodelan linier dan non-linier dapat menjadi strategi yang baik digunakan dalam praktek. Melalui kombinasi model yang berbeda, aspek yang berbeda dari pola data

dapat dikenali. Berdasarkan hal tersebut, cukup beralasan untuk membagi data runtun waktu atas struktur autokorelasi linier  $L_t$  dan struktur komponen non-linier  $N_t$ , dengan persamaan 34.

$$y_t = L_t + N_t \quad (34)$$

Kedua komponen diestimasi dari data di mana langkah pertama adalah pemodelan ARIMA untuk estimasi komponen linier. Selanjutnya, nilai kesalahan dari model tersebut dihitung dengan persamaan 35, di mana  $e_t$  hanya mengandung komponen non-linier.  $\hat{L}_t$  adalah hasil estimasi berdasarkan model ARIMA.

$$e_t = y_t - \hat{L}_t \quad (35)$$

Nilai kesalahan  $e_t$  cukup penting untuk dianalisis apakah pendekatan linier telah sesuai untuk data runtun waktu. Apabila nilai  $e_t$  masih memiliki struktur pola tertentu, maka model ARIMA belum sesuai. Selanjutnya  $e_t$  dimodelkan dengan ANN dengan  $n$  input menggunakan persamaan 36. Hasil akhir prediksi merupakan gabungan dari estimasi linier dan non linier, sesuai persamaan 37.

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon_t \quad (36)$$

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (37)$$

## 2.2.8 Data sintentik dan simulasi

Untuk menguji kemampuan model dalam melakukan seleksi variabel prediktor yang relevan dan model penduga data *spatio-temporal*, diperlukan data uji sintetik yang sudah diketahui perilaku, fungsi dan hubungannya dengan variabel respon. Berikut adalah prosedur untuk membangkitkan data sintetik dari curah hujan dan cuaca harian yang bersifat *spatio-temporal* [128], [129].

1. Membangkitkan data kejadian hujan.

1.a. Menghitung semua parameter dan kondisi awal.

1.a.1. Menghitung Matriks Korelasi Pearson curah hujan  $X$  dari lokasi  $i$  dan  $j$ .

$$\rho_{ij} = \frac{1}{n} \frac{\sum_{t=1}^n (x_i - \mu_i)(x_j - \mu_j)}{\sigma_i \sigma_j} \quad (38)$$

1.a.2. Menghitung N-jarak korelasi Euclidean dari 2 lokasi yang paling dependen dan urutkan secara *ascending*.

$$G_i = \sum_{j=1}^{n-1} (1 - |\rho_{ij}|) \quad (39)$$

1.a.3. Menghitung Probabilitas Transisi Markov dari 2 state yang orthogonal untuk lokasi  $i$  pada saat ke- $t$ .

$$\begin{cases} P_{1|0,0|0} = Pr[X_{it}[X_{jt} = 0 \wedge X_{kt} = 0|X_{i(t-1)} = 0] \\ P_{1|0,0|1} = Pr[X_{it}[X_{jt} = 0 \wedge X_{kt} = 0|X_{i(t-1)} = 1] \\ P_{1|0,1|0} = Pr[X_{it}[X_{jt} = 0 \wedge X_{kt} = 1|X_{i(t-1)} = 0] \\ P_{1|0,1|1} = Pr[X_{it}[X_{jt} = 0 \wedge X_{kt} = 1|X_{i(t-1)} = 1] \\ P_{1|1,0|0} = Pr[X_{it}[X_{jt} = 1 \wedge X_{kt} = 0|X_{i(t-1)} = 0] \\ P_{1|1,0|1} = Pr[X_{it}[X_{jt} = 1 \wedge X_{kt} = 0|X_{i(t-1)} = 1] \\ P_{1|1,1|0} = Pr[X_{it}[X_{jt} = 1 \wedge X_{kt} = 1|X_{i(t-1)} = 0] \\ P_{1|1,1|1} = Pr[X_{it}[X_{jt} = 1 \wedge X_{kt} = 1|X_{i(t-1)} = 1] \end{cases} \quad (40)$$

1.a.4. Bangkitkan jumlah hari dari kejadian hujan per bulan untuk setiap lokasi.

$$\hat{T}_m = \bar{T}_m \pm \sigma_m r_{nom}^{\wedge} \quad (41)$$

1.b. Membangkitkan korelasi spasial-temporal dari data kejadian hujan.

1.b.1. Bangkitkan data kejadian hujan dari 2 lokasi utama.

$$\rho_{ij} = f\left(\frac{X_{it=0 \wedge jt=0}}{n}, \frac{X_{it=1 \wedge jt=0}}{n}, \frac{X_{it=0 \wedge jt=1}}{n}, \frac{X_{it=1 \wedge jt=1}}{n}\right) \quad (42)$$

1.b.2. Gunakan probabilitas transisi markov untuk membangkitkan data kejadian hujan dari lokasi tersisa.

2. Membangkitkan data curah hujan.

2.a. Korelasi Person: Menghitung korelasi person setelah transformasi semua curah hujan harian dari distribusi Gamma kepada distribusi Gaussian.

2.b. Menghitung parameter distribusi Gamma ( $\alpha$ ,  $\beta$ ).

2.c. Membangkitkan data curah hujan untuk setiap lokasi.

$$\hat{R}_m = r_{gam}^* \beta_m \ln\{\Gamma(\alpha_m)\} \quad (43)$$

## 2.2.9 Ukuran kinerja model

Kinerja model penduga data runtun waktu  $Z_t$ , dapat menggunakan ukuran *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), Koefisien Determinan ( $R^2$ ) dan korelasi ( $r$ ) [130].

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Z_t - \hat{Z}_t)^2}{n}} \quad (44)$$



$$MAE = \frac{\sum_{t=1}^n |Z_t - \hat{Z}_t|}{n} \quad (45)$$

$$R_{sq} = 1 - \frac{\sum_{t=1}^n (Z_t - \hat{Z}_t)^2}{\sum_{t=1}^n (Z_t - \bar{Z})^2} \quad (46)$$

$$r = \sqrt{R_{sq}} \quad (47)$$

Untuk menetapkan nilai optimal dari *trade-off* antara RMSE dan jumlah prediktor  $p$  pada proses training (*in-sample*), digunakan *corrected AIC* (*Akaike Information Criterion*) dengan persamaan 48, AIC diperoleh dari persamaan 49 [5], [123].

$$AIC_{corrected} = AIC + \frac{2(p+1)(p+2)}{n-p} \quad (48)$$

$$AIC = -n \cdot \log \frac{\sum_{j=1}^n (y_j - \hat{y}_j(k))^2}{n} + 2(p+1) \quad (49)$$



## BAB 3

### METODE PENELITIAN

Sejumlah langkah dilakukan untuk membangun model penduga curah hujan Zona Musim dengan studi kasus wilayah Jawa Timur. Pada Gambar 3.1, dijelaskan diagram alir untuk pemodelan sebagai berikut:

a. Pengumpulan data.

Data curah hujan, suhu min/mean/maks, kelembaban udara, lama penyinaran matahari, rata-rata kecepatan angin, dan kecepatan angin maksimum diperoleh dari BMKG (<http://dataonline.bmkg.go.id/>). Sedangkan untuk data indeks iklim harian/bulanan, yaitu suhu permukaan laut yang meliputi Nino1.2, Nino2, Nino3, Nino3.4, Nino4 dan DMI diperoleh dari *The Royal Netherlands Meteorological Institute* (KNMI) secara online pada <http://climexp.knmi.nl/>. Periode data yang digunakan pada penelitian ini adalah Januari 1981–Desember 2016, baik harian maupun bulanan. Sebagai studi kasus adalah data cuaca dari 3 stasiun BMKG di wilayah Surabaya (Juanda, Perak I, Perak II) dan 2 stasiun BMKG di wilayah Malang (Karang Ploso, Karangtretes).

b. Tahap interpolasi dan agregasi.

Dilakukan interpolasi dari data yang tidak tersedia (*missing value*) maupun pencilan (*outliers*), khususnya data observasi cuaca, di mana *missing value* dan *outliers* antara 4,7% - 16,3%. Untuk pemodelan data runtun waktu bulanan, pengamatan cuaca harian dilakukan agregasi menjadi data bulanan.

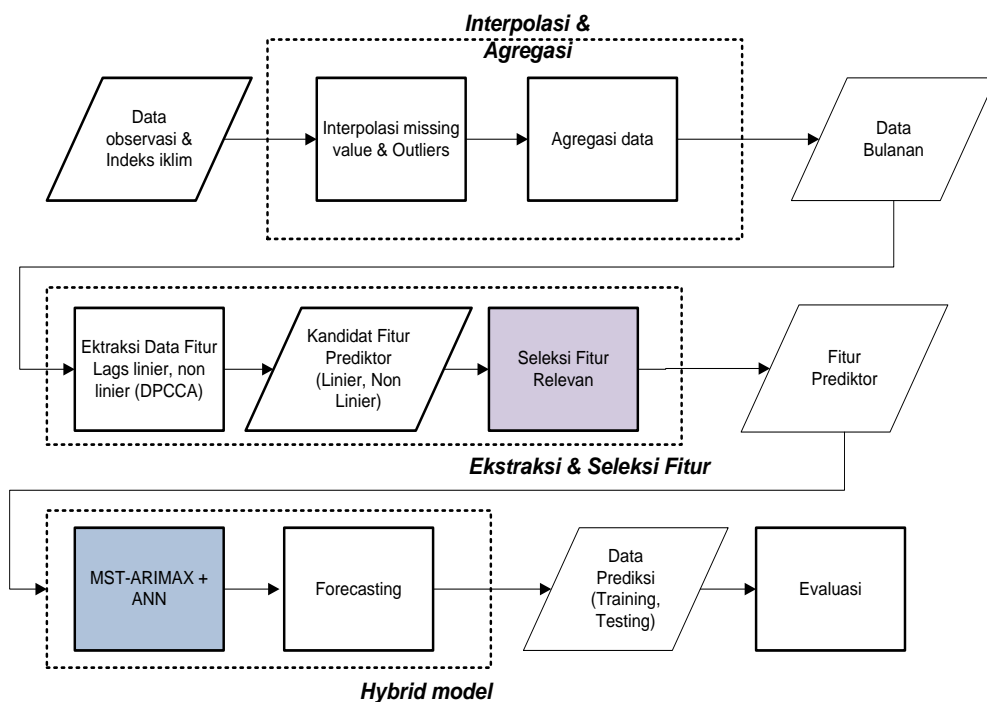
c. Tahap ekstraksi dan seleksi fitur (pra-proses).

Pemodelan dan prediksi curah hujan merupakan sistem yang kompleks dimana melibatkan interaksi berbagai faktor, yaitu *time-lags*, lokasi stasiun cuaca dan pengaruh iklim global juga diperhitungkan dalam pemodelan. Pada tahap ini, dilakukan ekstraksi fitur pengaruh *time-lags* terhadap curah hujan dengan pola relasi linier dan non-linier dan seleksi prediktor yang relevan sebagai penduga curah hujan.

d. Membangun model *hybrid* ST-ARX dan ANN.

Model ini merupakan pengembangan dari prediksi data runtun waktu multivariat yang memiliki pola hubungan linier dan non linier antara variabel

prediktor dengan variabel respon. Pada *Spatio-Temporal Autoregressive Integrated Moving-Average with eXogenous Variables* (ST-ARX), hubungan antar lokasi stasiun cuaca dan waktu diperhitungkan dalam model dan mengasumsikan bahwa hubungan antar variabel bersifat linier. Sedangkan untuk pola hubungan yang bersifat non-linier, digunakan metode *Artificial Neural Networks* (ANN) dengan input variabel prediktor non-linier yang telah diidentifikasi pada tahap (c) dan residual dari ST-ARX menjadi target dari ANN. Prediksi/peramalan merupakan nilai gabungan dari hasil ST-ARX dan ANN.



Gambar 3.1 Diagram alir model penduga curah hujan

- e. Prediksi (data training) dan peramalan (data testing).

Model *hybrid* diterapkan pada data training dan data testing untuk mengetahui kinerja dari model dalam melakukan peramalan curah hujan. Proporsi antara data training (kalibrasi) dan data testing (validasi) ditentukan berdasarkan persentase. Peramalan dapat dilakukan untuk satu periode ke depan (*one-step ahead*) maupun sejumlah periode ke depan (*multi-step ahead*), baik secara statis maupun dinamis.

- f. Tahap evaluasi

Evaluasi kinerja model diukur melalui nilai RMSE, MAE, R-square dan Korelasi ( $r$ ) antara nilai aktual dan nilai prediksi. Model memiliki kinerja tinggi

apabila memiliki nilai RMSE dan MAE yang rendah serta R-square dan  $r$  yang tinggi. Analisis terhadap nilai residual juga dilakukan untuk menjamin bahwa telah memenuhi *i.i.d.n* ( $0, \sigma^2$ ). Uji coba model dilakukan melalui beberapa pilihan, yaitu mengevaluasi kinerja model untuk ST-ARX, model ANN dan model *hybrid*. Skenario lainnya adalah uji coba tanpa menggunakan ekstraksi fitur maupun seleksi fitur untuk mengevaluasi efektifitas dari tahapan ini. Jumlah prediktor yang optimal di mana model menghasilkan kinerja terbaik diukur dengan *corrected AIC* pada proses training.

### 3.1 Tahap Interpolasi dan Agregasi

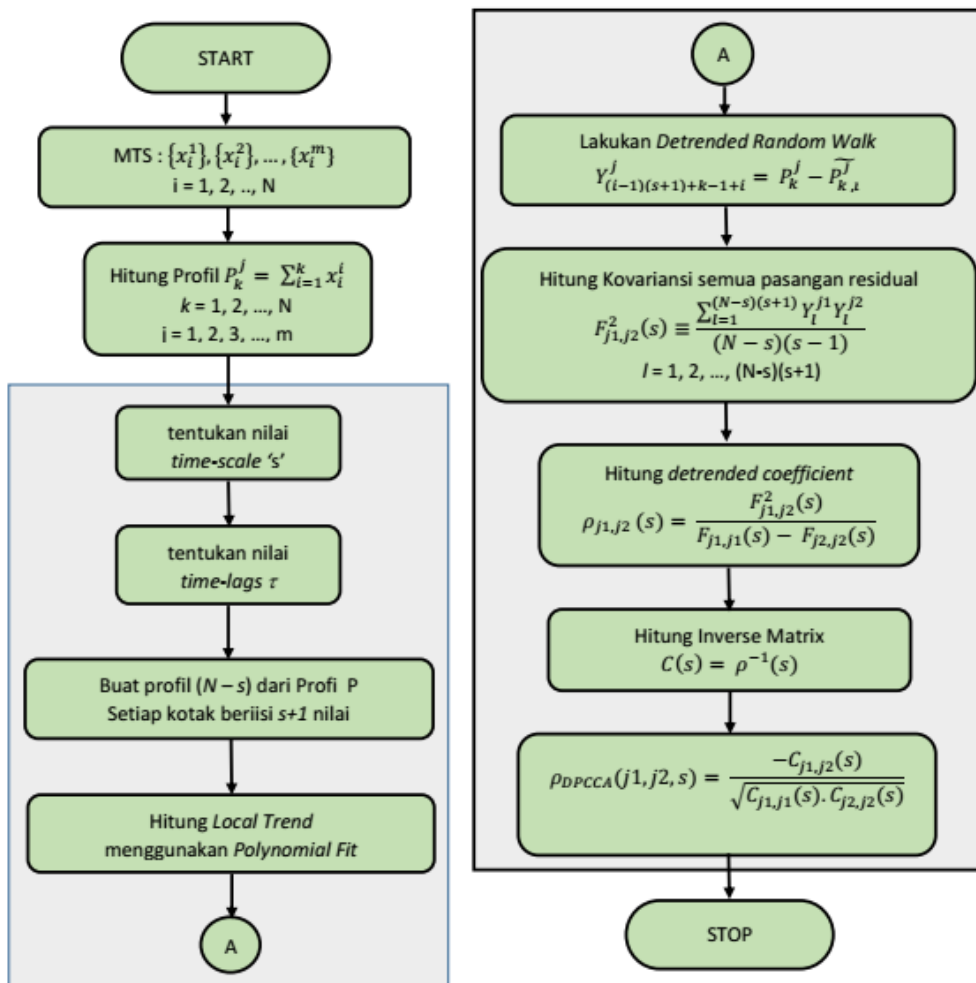
Tahap interpolasi digunakan untuk melengkapi *missing value* dari data meteorologi BMKG (data lokal). *Missing-value* diakibatkan kerusakan/terganggunya peralatan sehingga tidak teramati. Ada berbagai teknik interpolasi yang dapat digunakan untuk melengkapi *missing-value*, diantaranya adalah menggunakan nilai rata-rata. Sedangkan agregasi digunakan untuk mengubah data harian menjadi data bulanan. Berikut adalah agregasi untuk sejumlah data pengamatan lokal harian.

- a. Curah hujan bulanan: jumlah curah hujan harian pada satu bulan tertentu.
- b. Suhu udara minimum bulanan: nilai minimum dari data suhu minimum dalam satu bulan.
- c. Suhu udara maksimum bulanan: nilai maksimum dari data suhu maksimum dalam satu bulan.
- d. Suhu udara rata-rata bulanan: nilai rata-rata dari data suhu rata-rata dalam satu bulan.
- e. Kelembaban udara bulanan: nilai rata-rata dari data kelembaban udara dalam satu bulan.
- f. Kecepatan angin bulanan: nilai rata-rata dari data kecepatan angin harian dalam satu bulan.
- g. Lama penyinaran bulanan: nilai rata-rata dari data lama penyinaran harian dalam satu bulan.
- h. Kecepatan angin maksimum bulanan: nilai maksimum dari data kecepatan angin maksimum harian dalam satu bulan.

### 3.2 Tahap Ekstraksi dan Seleksi fitur

#### 3.2.1 Ekstraksi fitur tenggang waktu (time-lags)

Untuk mendapatkan himpunan kandidat fitur *time-lags* yang terkait dengan curah hujan dengan mempertimbangkan pengaruh lokasi stasiun cuaca (*spatial*), digunakan metode DPCCA [91] sesuai yang dijelaskan pada bagian 2.2.3. Langkah-langkah untuk memperoleh kandidat fitur *time-lags* dijelaskan pada Gambar 3.2.



Gambar 3.2 Flow-chart untuk ekstraksi fitur *time-lags* dengan DPCCA [91]

Data MTS dibentuk sebagai *random-walk*  $P_k^j$  dan kemudian dibuat *profil*  $P$  sebanyak  $(N-s)$ . Pola trend dihilangkan dengan model *Polynomial* sehingga tersisa runtun waktu residual. Nilai residual setiap MTS inilah yang dihitung nilai

kovariansi dan korelasi silang sehingga didapatkan matriks DPCCA. DPCCA yang optimal bergantung pada pemilihan skala waktu  $s$  dan *time-lags*  $\tau$ .

### 3.2.2 Tahap seleksi fitur relevan dan penghapusan redundan dengan PCSUF

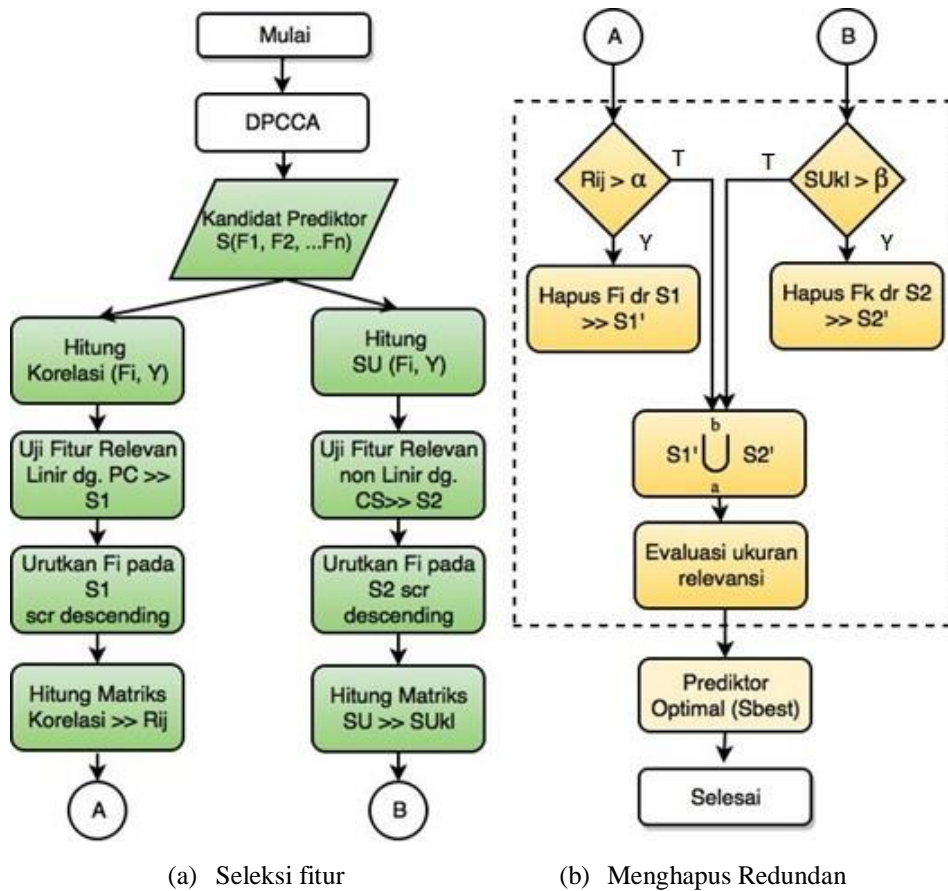
Zhang [49] dalam penelitiannya menyimpulkan bahwa tahap pra-proses sebelum pemodelan data runtun waktu, khususnya di bidang hidrologi, pra-proses terbukti meningkatkan kinerja prediksi. Pra-proses pada data runtun waktu di antaranya adalah berbasis pada transformasi, cara representasi data dan pra-proses berbasis ekstraksi/seleksi fitur [131].

Gambar 3.3 menunjukkan urutan proses seleksi prediktor (bagian a) dari input ekstraksi fitur menggunakan DPCCA. Kandidat prediktor yang relevan secara linier dipilih menggunakan korelasi pearson (PC), sedangkan prediktor non linier, digunakan *Symmetrical Uncertainty* (SU). Dari hasil keduanya, yaitu matriks  $R_{ij}$  dan matriks  $SU_{kl}$ , subset prediktor diseleksi kembali dengan menggunakan perbandingan nilai korelasi dan nilai SU terhadap nilai *therhold*  $\alpha$  dan  $\beta$  untuk menghapus fitur redundan (bagian b).

Himpunan prediktor optimal diperoleh melalui iterasi dari kombinasi nilai  $\alpha$  dan  $\beta$ . Gabungan 2 metode ini disebut PCSUF yang dikembangkan berdasarkan penelitian dari Jiang[94] dan Lei Yu[98]. Fungsi evaluasi( $\pi$ ) dari subset prediktor optimal menggunakan persamaan 50 yang merupakan adaptasi dari fungsi *merit value* [132]. Fungsi ini menyatakan bahwa nilai relevansi antara variabel respon  $Y$  dengan prediktor  $X_i$  sebanyak  $k$  dibagi relevansi antar variabel prediktor.

$$\pi_{X_{i,j},Y} = \frac{k(\overline{SU})_{Y,X_i}}{\sqrt{k + k(k-1)(\overline{SU})_{X_iX_j}}} \quad (50)$$

Detil dari proses seleksi fitur dan penghapusan fitur redundan menggunakan PCSUF dijelaskan pada *pseudocode* berikut.



Gambar 3.3 Seleksi fitur relevan dan hapus fitur redundan [133]

Input:  $S(F_1, F_2, \dots, F_n, Y)$  //data set with  $n$  features, one response variable

$\alpha, \beta$  // a predefined threshold

output:  $S_{best}$

```

1  Begin
2  For  $i = 1$  to  $n$  do begin
3      Calculate  $Corr_{i,Y}$  for  $F_i$ ;
4      If  $(Corr_{i,Y} \geq \text{value of critical table PC})$ 
5          append  $F_i$  to  $S1_{list}$ ;
6  End
7  For  $i = 1$  to  $n$  do begin
8      Calculate  $SU_{i,Y}$  for  $F_i$ ;
9      if  $(SU_{i,Y} \geq \text{value of critical table Chi-Square})$ 
10         append  $F_i$  to  $S2_{list}$ ;
11  Order  $S1_{list}$  and  $S2_{list}$  in descending
12   $F_j = \text{getFirstElement at } (S1_{list})$ 
13  do begin
14       $F_i = \text{getNextElement}(S1_{list}, F_j)$ 
15      if  $(F_i \neq \text{Null})$ 
16          do begin
17              If  $(Corr_{i,j} \geq \alpha)$ 
18                  remove  $F_i$  from  $S1_{list}$ 
19               $F_i = \text{getNextElement}(S1_{list}, F_i)$ 
20          End until  $(F_i == \text{Null})$ 
21       $F_j = \text{getNextElement}(S1_{list}, F_j)$ 

```



```

22 End until (Fj==Null)
23 Fj= getFirstElement at (S2'list)
24 do begin
25   Fi=getNextElement(S1'list, Fj)
26   if (Fi <> Null)
27   do begin
28     If (SUij ≥ β)
29       remove Fi from S2'list
30       Fi=getNextElement(S2'list, Fj)
31   End until (Fi==Null)
32   Fj=getNextElement(S2'list, Fj)
33 End until (Fj==Null)
34 Sbest=S1'list ∩ S2'list

```

### 3.3 Pemodelan hybrid ST-ARX dan ANN

Narendra [125] dan Zhang [127] dalam penelitiannya menyatakan bahwa dalam kasus riil peramalan data runtun waktu, sangat jarang ditemukan pola data yang murni linier atau non linier. Pada umumnya adalah gabungan di antara kedua pola tersebut. Oleh karenanya, model ARIMA atau ANN tidak cukup mampu memodelkan atau meramalkan data runtun waktu yang demikian. Hal ini karena ARIMA memiliki kekurangan tidak bisa mengenali pola non linier, sedangkan ANN tidak bisa mengenali pola hubungan yang bersifat linier.

Kedua penelitian model *hybrid* tersebut diterapkan pada data UTS untuk berbagai data sintetik dan data riil. Banihabib [15] dan Jain [38] telah menerapkan model *hybrid* ARIMA-ANN pada bidang meteorologi dan terbukti meningkatkan nilai akurasi dari penggunaan model parsial. Oleh karenanya dalam penelitian ini dikembangkan model hybrid ST-ARX dan ANN. Hal ini sejalan dengan kesimpulan dari penelitian Makridakis dan Hibbon [112] yang menyatakan bahwa kombinasi sejumlah metode lebih baik dibanding metode tunggal dalam model peramalan data runtun waktu.

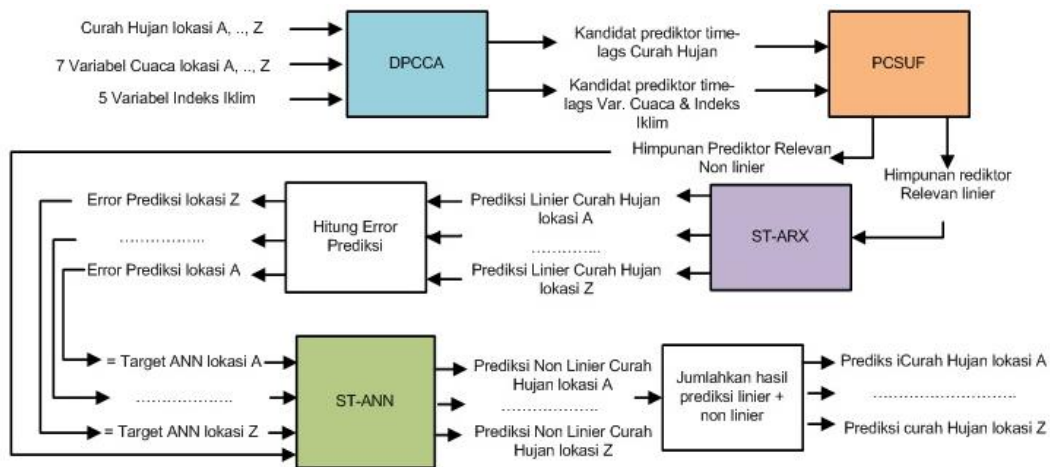
ST-ARX merupakan pengembangan dari model STARIMA dengan memperhitungkan variabel lokasi stasiun cuaca (spasial) dan menggunakan variabel eksternal sebagai prediktor, sehingga modelnya disebut sebagai *Spatio-Temporal ARIMA with eXogenous variables*. Persamaan 51 merupakan model ST-ARX [134] yang merupakan bentuk khusus ST-ARIMAX. Model ST-ARX digunakan karena untuk komponen *Moving-Average* (MA) akan dimodelkan sebagai ST-ANN pada tahap berikutnya.

$$y_{it} = \sum_{k=1}^p \sum_{i=0}^{\alpha_k} \theta_{kl} \sum_{j=1}^N w_{ij}^{(l)} y_{j;t-k} + \sum_{k=1}^r \sum_{i=0}^{\delta_k} \phi_{kl} \sum_{j=1}^N \omega_{ij}^{(l)} x_{j(t-k)} + a_{it} \quad (51)$$

- $y_{it}$  = variabel respon pada lokasi  $i$ , waktu ke- $t$   
 $p$  = order dari AR untuk variabel respon ( $y$ ).  
 $\alpha_k$  = order spasial dari AR ke- $k$  untuk variabel respon ( $y$ ).  
 $\theta_{kl}$  = parameter AR dari variabel respon ( $y$ ).  
 $w_{ij}^{(l)}$  = bobot interaksi antara  $y_{it}$  dan  $y_{j;t-k}$   
 $N$  = jumlah unit spasial  
 $r$  = order dari AR untuk variabel prediktor ( $x_j$ )  
 $\delta_k$  = order spasial dari AR ke- $k$  untuk variabel prediktor ( $x$ ).  
 $\phi_{kl}$  = parameter AR untuk variabel prediktor ( $x_j$ )  
 $\omega_{ij}^{(l)}$  = bobot interaksi antara  $y_{it}$  dan variabel prediktor ( $x$ )  
 $a_{it}$  = residual model

Pengembangan model penduga curah hujan melalui pra-proses (DPCCA dan PCSUF), model *hybrid* ST-ARX dan ANN dijelaskan pada Gambar 3.4. Input untuk model ST-ARX adalah sejumlah variabel prediktor linier relevan ( $X_{i,t-k}$ ) dari variabel prediktor cuaca, indeks iklim dan pengaruh dari curah hujan periode sebelumnya  $Y_{i,t-k}$  yang telah diidentifikasi menggunakan PCSUF dengan variabel respon adalah curah hujan pada waktu ke- $t$ .

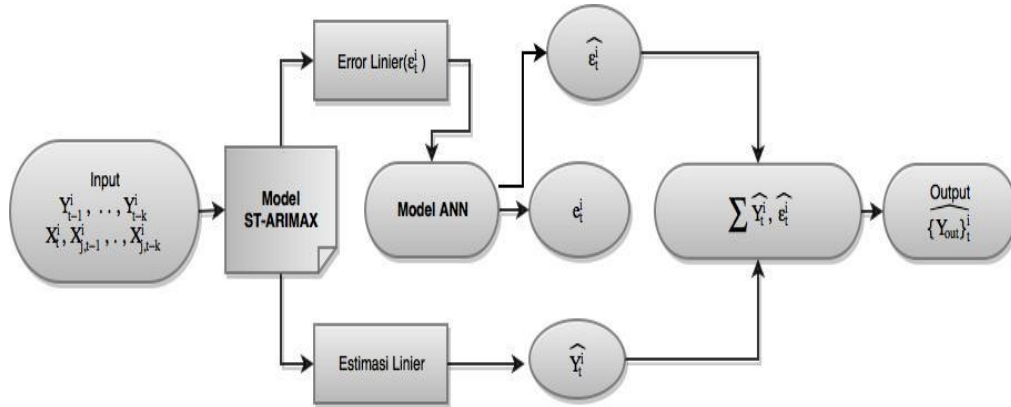
Luaran dari ST-ARX adalah prediksi linier dari curah hujan untuk sejumlah lokasi stasiun cuaca dan nilai *error* dari model ST-ARX. Nilai *error* menjadi target dari pemodelan ANN dan sebagai input dari ANN adalah variabel cuaca dan indeks iklim yang berpola non-linier. Hasil dari ANN adalah prediksi atas nilai *error* ST-ARX yang juga merupakan prediksi curah hujan untuk komponen non linier.



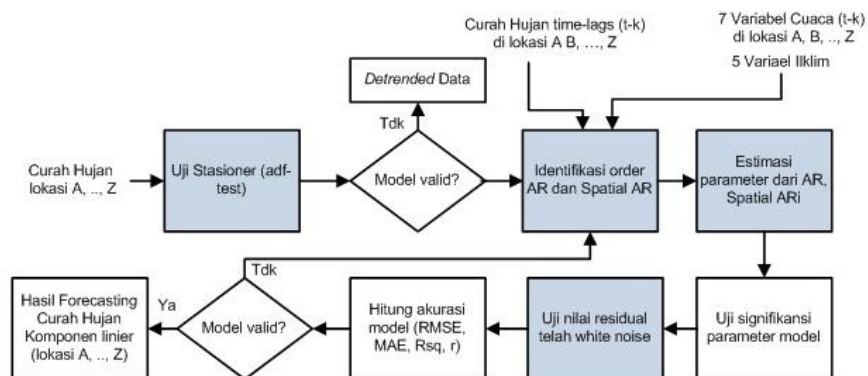
Gambar 3.4 Pra-proses, hybrid ST-ARX dan ANN

Luaran dari model *hybrid* ST-ARX-ANN adalah jumlahan dari luaran ST-ARX dan luaran ANN sehingga menjadi nilai prediksi curah hujan yang lengkap.

Gambar 3.5 adalah proses pemodelan *hybrid* STARX dan ANN. Hasil dari tahap ini adalah prediksi curah hujan dengan prediktor linier ( $Y'_{linier}$ ) dan kesalahan (*error*) dari model ( $E_{linier}$ ). Estimasi parameter model ( $\theta_{kl}$  dan  $\varphi_{kl}$ ) dapat dilakukan dengan metode *conditional least square*, *maximum likelihood function* [126], ataupun pilihan lainnya [130]. Sedangkan penentuan bobot  $w_{ij}^{(l)}$  ditentukan di awal pemodelan. Tahap untuk membangun model ST-ARX meliputi identifikasi stasioner dari variabel respon, identifikasi model, estimasi parameter dan evaluasi kinerja model [134] seperti yang dijelaskan pada Gambar 3.6. Model ST-ARX disebut valid apabila 4 kinerja pemodelan telah mencapai hasil yang optimal.



Gambar 3.5 Model *hybrid* STARX-ANN

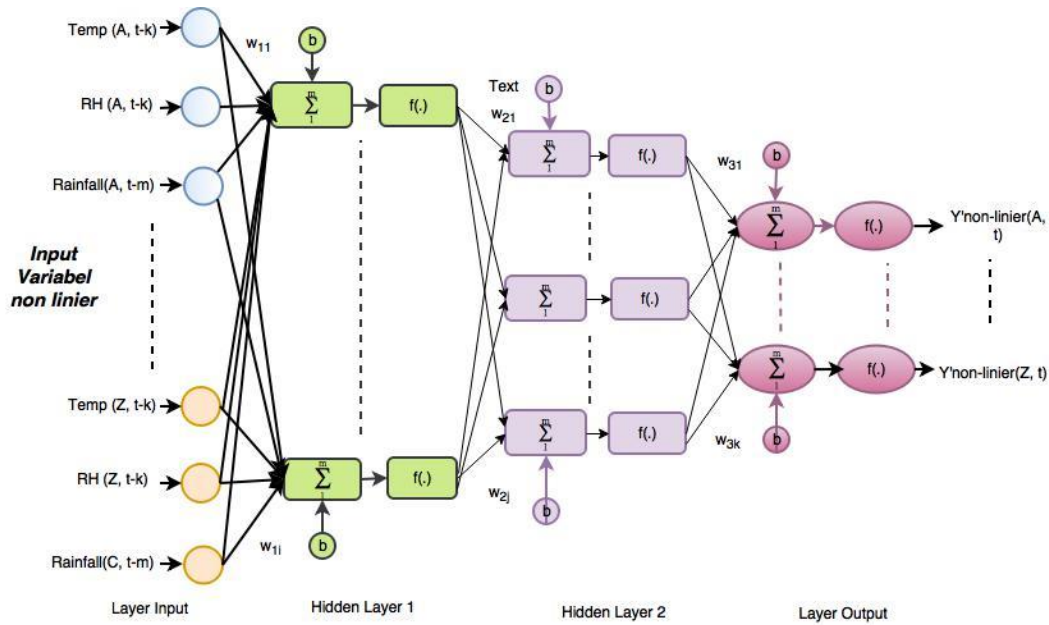


Gambar 3.6 Detil proses pemodelan ST-ARX

Selanjutnya  $E_{linier}$  menjadi target dari model ANN. Gambar 3.7 adalah detil pemodelan ANN. Input dari ANN adalah sejumlah variabel prediktor non linier berdasarkan lokasi stasiun cuaca dan dependensi tenggang waktu yang telah

diidentifikasi dengan menggunakan PCSUF. Oleh karenanya, model ANN tersebut dapat diistilahkan sebagai *Spatio-Temporal ANN* [108], [135].

Estimasi parameter (bobot) dapat menggunakan algoritma *Levenberg–Marquardt* [51], *back-propagation* [64], *Multi Layer Perceptron* [111]. Output dari ANN adalah prediksi curah hujan dengan diktor non linier ( $Y'_{\text{non-linier}}$ ). Arsitektur pada Gambar 3.7 terdiri atas 2 lapisan tersembunyi dengan proses training *Backpropagation*. Hasil keduanya digabungkan untuk memperoleh prediksi akhir ( $Y' = Y'_{\text{linier}} + Y'_{\text{non-linier}}$ ). [136].



Gambar 3.7 Spatio-temporal ANN

Proses training *Backpropagation* dapat dilakukan dengan cara menghapus sejumlah unit pemroses (*drop-out*) pada lapisan tersembunyi sehingga mengurangi kompleksitas, sesuai ilustrasi pada Gambar 3.8a [137]. Unit yang dihapus ditentukan secara random berdasarkan distribusi Bernoulli. Maka proses training untuk ANN dengan formula standar sesuai persamaan 52 dan persamaan 53 berubah menjadi persamaan 54-56 sesuai yang diilustrasikan pada(Gambar 3.8b.

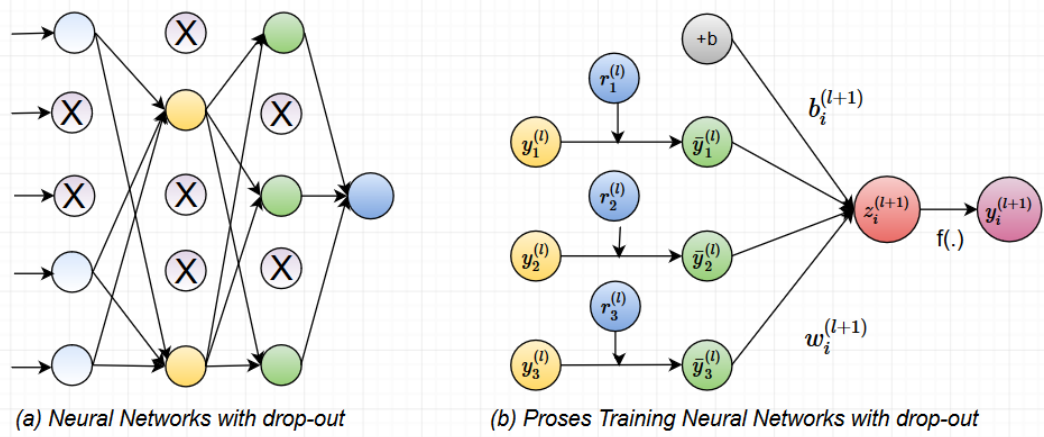
$$z_i^{(l+1)} = w_i^{(l+1)} y^l + b_i^{(l+1)} \quad (52)$$

$$y_i^{(l+1)} = f(z_i^{(l+1)}) \quad (53)$$

$$\tilde{y}^{(l)} = r^{(l)} * y^{(l)} \quad (54)$$

$$z_i^{(l+1)} = w_i^{(l+1)} \tilde{y}^{(l)} + b_i^{(l+1)} \quad (55)$$

$$y_i^{(l+1)} = f(z_i^{(l+1)}) \quad (56)$$



Gambar 3.8 Neural Networks with drop-out [137]

### 3.4 Evaluasi Kinerja

Uji coba untuk mengevaluasi kinerja model dalam meningkatkan efektifitas peramalan data runtun waktu dilakukan pada 2 studi kasus, yaitu data sintetis dan data kasus riil. Evaluasi atas pengaruh faktor-faktor tersebut terhadap efektifitas model didasarkan pada nilai RMSE, MAE,  $R^2$ ,  $r$  untuk proses training (kalibrasi) maupun validasi (testing), dan *corrected* AIC untuk proses training.

#### 3.4.1 Evaluasi model dengan data sintetis

Data sintetis dan simulasi yang dibangkitkan dengan fungsi seperti pada sub bab 2.2.8, akan dimodelkan dengan pendekatan *hybrid* ARX dan ANN dengan terlebih dahulu dilakukan seleksi fitur prediktor. Data sintetis telah diketahui mempunyai fungsi linier dan non linier, mengandung variabel tenggang waktu dan persamaannya telah diketahui.

Evaluasi uji coba dan evaluasinya didasarkan pada 3 faktor, yaitu:

- Identifikasi pola variabel prediktor (linier/non linier).
- Efektifitas prosedur seleksi fitur prediktor.
- Penggunaan model *hybrid* ARX dan ANN.

Tabel 3.1 menunjukkan kombinasi kategori dari setiap faktor untuk skenario evaluasi data sintetis. Diharapkan kombinasi pengujian nomor 9 akan menghasilkan kinerja model terbaik karena telah dilakukan identifikasi pola hubungan linier dan non linier, seleksi fitur prediktor dan penerapan model *hybrid*.

Tabel 3.1 Evaluasi untuk data sintetis

No. Uji	Pola		Seleksi Fitur		Model	
	Linier	Non Linier	Ya	Tdk	ARX	ANN
1	√		√		√	
2	√			√	√	
3	√		√		√	
4	√			√	√	
5		√	√			√
6		√		√		√
7		√	√			√
8		√		√		√
9	√	√	√		√	√
10	√	√		√	√	√

### 3.4.2 Evaluasi model dengan data riil

Uji coba untuk data meteorologi dan iklim, dievaluasi berdasarkan 3 (tiga) faktor dan pengaruhnya terhadap efektifitas model yang diusulkan. Hipotesis terhadap faktor-faktor tersebut adalah:

- Penggunaan prediktor lokal (cuaca) dan indeks iklim (global) akan meningkatkan akurasi prediksi.
- Metode PCSUF akan meningkatkan akurasi prediksi.
- Pendekatan spasial (interdependensi antar lokasi stasiun cuaca) akan meningkatkan akurasi prediksi.

Tabel 3.2 menunjukkan kombinasi kategori dari setiap faktor untuk skenario uji coba dari 3 data riil.

Tabel 3.2 Skenario evaluasi untuk model penduga curah hujan

No. Uji	Prediktor		seleksi fitur		Spatial	
	Lokal	Global	Ya	Tdk	Ya	Tdk
1	√	√			√	
2	√		√		√	
3	√	√				√
4	√		√			√
5		√		√		√
6			√	√		
7	√	√		√	√	
8	√		√	√	√	
9	√	√		√		√
10	√		√	√		√

### 3.5 Hasil-Hasil Penelitian Awal

Peneliti telah melakukan studi ekstraksi fitur dengan menggunakan CCF dan seleksi prediktor dengan PCSUF pada data meteorologi stasiun BMKG Juanda untuk curah hujan bulanan berdasarkan variabel prediktor lokal dan indeks iklim. Variabel lokal yang dimaksud adalah curah hujan, suhu minimum/maksimum/rata-rata, kelembaban udara, lama penyinaran, nino3, nino 4, nino34, SOI dan SPL. Data yang digunakan adalah periode Januari 2001 – Desember 2015 periode harian.

Hasil penelitian tersebut menunjukkan bahwa dengan metode ekstraksi fitur, semua variabel prediktor berkorelasi signifikan dengan variabel curah hujan dengan tenggang waktu ( $t-k$ ) tertentu. Demikian juga metode PCSUF dalam seleksi variabel prediktor, berhasil mengurangi jumlah prediktor sebesar 77,9% (68 menjadi 15 fitur prediktor) dan meningkatkan *merit-value* menjadi 2,25 kali bila tanpa dilakukan seleksi prediktor [138].

Peneliti juga telah melakukan studi perbandingan kinerja 3 model penduga curah hujan, yaitu metode ARIMA dan *Single-Input* dan *Multi-Input Transfer Function* (ARIMAX). Studi kasus yang digunakan adalah data curah hujan dari periode Januari 1981 – Desember 2016 dan 7 fitur prediktor lokal dan 5 prediktor global. Hasil dari studi ini menunjukkan bahwa model *Multi-Input Transfer Function* merupakan model terbaik dengan 4 variabel prediktor, yaitu temperatur rata-rata, temperatur maksimum, kelembaban udara dan lama penyinaran.

Sedangkan prediktor global tidak mempengaruhi curah hujan. Tingkat kesesuaian model terhadap data riil adalah 72% dengan nilai korelasi ( $r$ ) antara data riil dan data prediksi adalah 0.85 [120].

Studi awal terhadap karakteristik himpunan variabel prediktor dan curah hujan, juga telah dilakukan. Karakteristik yang dimaksud yaitu hubungan antara variabel lokal dan global terhadap curah hujan, bersifat linier/non linier dan pola data runtun waktu dari data tersebut, bersifat stasioner/non stasioner, serta pengaruh dari dependensi lokasi stasiun cuaca dan tenggang waktu (*spatio-temporal*) untuk data cuaca dan indeks iklim (jan 1981-des 2016).

Korelasi linier antara variabel prediktor dengan variabel curah hujan diuji dengan PC, sedangkan korelasi non linier diuji dengan *white-test*. Stasioneritas dari pola data runtun waktu diuji dengan menggunakan *adf test*, sedangkan pengaruh dari tenggang waktu diuji dengan CCF. Sedangkan dependensi antar variabel berbeda lokasi (*spatial*) diuji dengan PC. Pada Tabel 3.3 merupakan ringkasan dari hasil identifikasi dari karakteristik himpunan variabel tersebut.

Tabel 3.3 karakteristik variabel curah hujan, prediktor lokal dan global

Variabel	Korelasi dengan curah hujan		Pola data runtun waktu		Pengaruh spatio-temporal	
	Linier	non-linier	stasioner	non-stasioner	time-lags	lokasi
Curah hujan			√			√
Suhu min	√	√	√		√	√
Suhu maks			√		√	√
Suhu rata <sup>2</sup>		√	√		√	√
Kelembaban	√	√	√		√	√
Lama Penyinaran	√	√	√		√	√
Kec. Angina	√	√	√		√	√
Kec. angin maks	√		√		√	√
Nino1.2		√		√	√	
Nino3		√		√	√	
Nino4		√		√		
Nino34		√		√		
DMI		√		√		



## DAFTAR PUSTAKA

- [1] E. Makmur, Y. Koesmaryono, E. Aldrian, and A. H. Wigena, "Wet Season Onset Prediction Model Upon Rice Fields In Northern West Java Using Regional And Global Predictor," *J. Meteorol. dan Geofis.*, vol. 14, pp. 127–137, 2013.
- [2] Y. S. Swarinoto, Y. Koesmaryono, E. Aldrian, and A. H. Wigena, "Ensemble Prediction System Model For Monthly Rainfall Total Using Weighting Values (Case Of Indramayu District)," *J. Mkg*, vol. 13, no. 3, pp. 189–200, 2012.
- [3] S. Asadi, J. Shahrabi, P. Abbaszadeh, and S. Tabanmehr, "Neurocomputing A new hybrid artificial neural networks for rainfall – runoff process modeling," *Neurocomputing*, vol. 121, pp. 470–480, 2013.
- [4] Y. Zou, W. Xue, and S. Liu, "A case study of large-scale parallel I/O analysis and optimization for numerical weather prediction system," *Futur. Gener. Comput. Syst.*, vol. 37, pp. 378–389, 2014.
- [5] R. J. May, H. R. Maier, G. C. Dandy, and T. M. K. G. Fernando, "Non-linear variable selection for artificial neural networks using partial mutual information," *Environ. Model. Softw.*, vol. 23, pp. 1312–1326, 2008.
- [6] S. Bhattacharji, H. J. Neugebauer, J. Reitner, G. M. Friedman, and A. Seilacher, "Nonlinear Time Series Analysis in the Geosciences," in *Lecture Notes in Earth Sciences*, 2008, p. 390.
- [7] A. Sharma and M. K. Goyal, "Bayesian Network Model for Monthly Rainfall Forecast," *2015 IEEE Int. Conf. Res. Comput. Intell. Commun. Networks (ICRCICN), Kolkata, India*, pp. 241–246, 2015.
- [8] M. Han, W. Ren, and X. Liu, "Joint mutual information-based input variable selection for multivariate time series modeling," *Eng. Appl. Artif. Intell.*, vol. 37, pp. 250–257, 2015.
- [9] R. V. Ramana, B. Krishna, and S. R. Kumar, "Monthly Rainfall Prediction Using Wavelet Neural Network Analysis," *Water Resour. Manag.*, vol. 27, pp. 3697–3711, 2013.
- [10] J. Sánchez-Monedero, S. Salcedo-Sanz, P. A. Gutiérrez, C. Casanova-Mateo, and C. Hervás-Martínez, "Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-ordinal support vector classifier," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 199–207, 2014.
- [11] E. G. Ortiz-García, S. Salcedo-Sanz, and C. Casanova-Mateo, "Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data," *Atmos. Res.*, vol. 139, pp. 128–136, 2014.
- [12] R. Hashim *et al.*, "Selection of meteorological parameters affecting rainfall estimation using neuro-fuzzy computing methodology," *Atmos. Res.*, vol. 171, pp. 21–30, 2016.
- [13] F. J. Chang, L. C. Chang, C. W. Huang, and I. F. Kao, "Prediction of monthly regional groundwater levels through hybrid soft-computing techniques," *J. Hydrol.*, vol. 541, pp. 965–976, 2016.
- [14] N. Acharya, N. A. Shrivastava, B. K. Panigrahi, and U. C. Mohanty, "Development of an artificial neural network based multi-model ensemble

- to estimate the northeast monsoon rainfall over south peninsular India: An application of extreme learning machine,” *Clim. Dyn.*, vol. 43, no. 5–6, pp. 1303–1310, 2014.
- [15] M. E. Banihabib, A. Ahmadian, and F. S. Jamali, “Hybrid DARIMA-NARX model for forecasting long-term daily inflow to Dez reservoir using the North Atlantic Oscillation (NAO) and rainfall data,” *GeoResJ*, vol. 13, pp. 9–16, 2017.
  - [16] X. He, H. Guan, and J. Qin, “A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall,” *J. Hydrol.*, vol. 527, pp. 88–100, 2015.
  - [17] A. H. Setyaningrum and P. M. Swarinata, “Weather prediction application based on ANFIS (Adaptive neural fuzzy inference system) method in West Jakarta region,” *2014 Int. Conf. Cyber IT Serv. Manag. CITSM 2014*, pp. 113–118, 2014.
  - [18] A. Nugroho and B. H. Simanjuntak, “ARMA (Autoregressive Moving Average) Model for Prediction of Rainfall in Regency of Semarang - Central Java - Republic of Indonesia,” *Int. J. Comput. Sci. Issues*, vol. 11, no. 3, pp. 27–32, 2014.
  - [19] A. Nugroho, S. Hartati, and K. Mustofa, “Vector Autoregression (Var) Model for Rainfall Forecast and Isohyet Mapping in Semarang–Central Java–Indonesia,” ... *J. Adv. Comput. Sci.* ..., vol. 5, no. 11, pp. 44–49, 2014.
  - [20] H. Lee, “General Rainfall Patterns in Indonesia and the Potential Impacts of Local Seas on Rainfall Intensity,” *Water*, vol. 7, no. 4, pp. 1751–1769, 2015.
  - [21] S. Hardwinarto and M. Aipassa, “Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggara Station, East Kalimantan - Indonesia,” *Procedia Comput. Sci.*, vol. 59, no. Iccsci, pp. 142–151, 2015.
  - [22] F. Nhita, S. Annisa, and S. Kinasih, “Comparative Study of Grammatical Evolution and Adaptive Neuro-Fuzzy Inference System on Rainfall Forecasting in Bandung,” in *2015 3rd International Conference on Information and Communication Technology (ICoICT) A.*, 2015, pp. 6–10.
  - [23] M. N. Nur’utami and R. Hidayat, “Influences of IOD and ENSO to Indonesian Rainfall Variability: Role of Atmosphere-ocean Interaction in the Indo-pacific Sector,” *Procedia Environ. Sci.*, vol. 33, pp. 196–203, 2016.
  - [24] D. Munandar, “Optimation Weather Parameters Influencing Rainfall Prediction Using Adaptive Network-Based Fuzzy Inference System (ANFIS) and Linear Regression,” in *2015 International Conference on Data and Software Engineering OptimizationWeather*, 2015, pp. 154–159.
  - [25] S. C. Nisak, “Seemingly Unrelated Regression Approach for GSTARIMA Model to Forecast Rain Fall Data in Malang Southern Region Districts,” *Cauchy - J. Pure Appl. Math.*, vol. 4, no. 2, pp. 57–64, 2016.
  - [26] J. Toggweiler and R. Key, “Ocean circulation: Thermohaline circulation,” *Encycl. Atmos. Sci.*, vol. 4, no. December 2007, p. 1549–1555., 2001.
  - [27] J.-I. Hamada, M. D. Yamanaka, J. Matsumoto, S. Fukao, P. A. Winarso, and T. Sribimawati, “Spatial and Temporal Variations of the Rainy Season over Indonesia and their Link to ENSO,” *J. Meteorol. Soc. Japan*, vol. 80, no. 2, pp. 285–310, 2002.

- [28] E. Aldrian and R. Dwi Susanto, "Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature," *Int. J. Climatol.*, vol. 23, no. 12, pp. 1435–1452, 2003.
- [29] H. H. Hendon, "Indonesian rainfall variability: Impacts of ENSO and local air-sea interaction," *J. Clim.*, vol. 16, no. 11, pp. 1775–1790, 2003.
- [30] Supari, "SPATIOTEMPORAL CHARACTERISTICS OF EXTREME RAINFALL EVENTS OVER JAVA ISLAND , Case: East Java Province," 2012.
- [31] E. Aldrian, *Meteorologi Laut Indonesia*, no. November. Jakarta: Badan Meteorologi dan Geofisika, 2008.
- [32] B. Widjanarko Otok, "Development of Rainfall Forecasting Model in Indonesia by using ASTAR, Transfer Function, and ARIMA Methods," *Eur. J. Sci. Res.*, vol. 38, no. 3, pp. 1450–216, 2009.
- [33] D. J. Vimont, D. S. Battisti, and R. L. Naylor, "Downscaling Indonesian precipitation using large-scale meteorological fields," *Int. J. Climatol.*, vol. 30, no. 11, pp. 1706–1722, 2010.
- [34] M. V. Fernandes, A. M. Schmidt, and H. S. Migon, "Modelling zero-inflated spatio-temporal processes," *Stat. Modelling*, vol. 9, no. 1, pp. 3–25, 2009.
- [35] M. Fuentes, B. Reich, and G. Lee, "Spatial-temporal mesoscale modeling of rainfall intensity using gage and radar data," *Ann. Appl. Stat.*, vol. 2, no. 4, pp. 1148–1169, 2008.
- [36] J. H. Faghmous and V. Kumar, "Spatio-Temporal Data Mining for Climate Data : Advances , Challenges , and Opportunities," *Data Min. Knowl. Discov. Big Data Vol. 1*, pp. 83–116, 2013.
- [37] Lilies Handayani, "Statistical Downscaling Dengan Model Aditif Terampat Untuk Pendugaan Curah Hujan Ekstrem," *Thesis Sekol. Pasca Sarj. IPB*, 2014.
- [38] A. Jain and A. M. Kumar, "Hybrid neural network models for hydrologic time series forecasting," *Appl. Soft Comput. J.*, vol. 7, no. 2, pp. 585–592, 2007.
- [39] A. M. Kalteh, "Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform," *Comput. Geosci.*, vol. 54, pp. 1–8, 2013.
- [40] H. Badrzadeh, R. Sarukkalige, and A. W. Jayawardena, "Hourly runoff forecasting for flood risk management: Application of various computational intelligence models," *J. Hydrol.*, vol. 529, pp. 1633–1643, 2015.
- [41] R. C. Deo and M. Şahin, "Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia," *Atmos. Res.*, vol. 153, pp. 512–525, 2015.
- [42] H. Memarian, M. Pourreza Bilondi, and M. Rezaei, "Drought prediction using co-active neuro-fuzzy inference system, validation, and uncertainty analysis (case study: Birjand, Iran)," *Theor. Appl. Climatol.*, vol. 125, no. 3–4, pp. 541–554, 2016.
- [43] R. C. Deo, P. Samui, and D. Kim, "Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models," *Stoch. Environ. Res. Risk Assess.*, vol. 30, no. 6, pp. 1769–1784, 2016.

- [44] M. Tiwari, J. Adamowski, and K. Adamowski, "Water demand forecasting using extreme learning machines," *J. Water L. Dev.*, vol. 28, no. 1, 2016.
- [45] J. Du, Y. Liu, Y. Yu, and W. Yan, "A prediction of precipitation data based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) algorithms," *Algorithms*, vol. 10, no. 2, 2017.
- [46] F. O. Hocaoglu and F. Serttas, "A novel hybrid (Mycielski-Markov) model for hourly solar radiation forecasting," *Renew. Energy*, vol. 108, pp. 635–643, 2015.
- [47] J. Marohasy and J. Abbot, "Assessing the quality of eight different maximum temperature time series as inputs when using artificial neural networks to forecast monthly rainfall at Cape Otway, Australia," *Atmos. Res.*, vol. 166, pp. 141–149, 2015.
- [48] M. Han and W. Ren, "Global mutual information-based feature selection approach using single-objective and multi-objective optimization," *Neurocomputing*, vol. 168, pp. 47–54, 2015.
- [49] X. Zhang, Y. Peng, C. Zhang, and B. Wang, "Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences," *J. Hydrol.*, vol. 530, pp. 137–152, 2015.
- [50] J. Anderson, H. Van Den Dool, A. Barnston, W. Chen, W. Stern, and J. Ploshay, "Present-Day Capabilities of Numerical and Statistical Models for Atmospheric Extratropical Seasonal Simulation and Prediction," *Bull. Am. Meteorol. Soc.*, vol. 80, no. 7, pp. 1349–1361, 1999.
- [51] V. Nourani, Ö. Kisi, and M. Komasi, "Two hybrid Artificial Intelligence approaches for modeling rainfall-runoff process," *J. Hydrol.*, vol. 402, no. 1–2, pp. 41–59, 2011.
- [52] Z. M. Yaseen, A. El-shafie, O. Jaafar, H. A. Afan, and K. N. Sayl, "Artificial intelligence based models for stream-flow forecasting: 2000-2015," *J. Hydrol.*, vol. 530, pp. 829–844, 2015.
- [53] X. Wang and M. Han, "Improved extreme learning machine for multivariate time series online sequential prediction," *Eng. Appl. Artif. Intell.*, vol. 40, pp. 28–36, 2015.
- [54] A. Talei, L. H. C. Chua, C. Quek, and P. E. Jansson, "Runoff forecasting using a Takagi-Sugeno neuro-fuzzy model with online learning," *J. Hydrol.*, vol. 488, pp. 17–32, 2013.
- [55] M. Shoaib, A. Y. Shamseldin, B. W. Melville, and M. Muneer, "Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach," *J. Hydrol.*, vol. 527, pp. 326–344, 2015.
- [56] E. G. Ortiz-garcía, S. Salcedo-sanz, and C. Casanova-mateo, "Accurate precipitation prediction with support vector classifiers : A study including novel predictive variables and observational data," *Atmos. Res.*, vol. 139, pp. 128–136, 2014.
- [57] O. Kisi and M. Cimen, "Precipitation forecasting by using wavelet-support vector machine conjunction model," *Eng. Appl. Artif. Intell.*, vol. 25, no. 4, pp. 783–792, 2012.
- [58] M. Kühnlein, T. Appelhans, B. Thies, and T. Nauss, "Improving the accuracy of rainfall rates from optical satellite sensors with machine learning

- A random forests-based approach applied to MSG SEVIRI,” *Remote Sens. Environ.*, vol. 141, pp. 129–143, 2014.
- [59] C. C. F. Plouffe, C. Robertson, and L. Chandrapala, “Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka,” *Environ. Model. Softw.*, vol. 67, pp. 57–71, 2015.
- [60] J. Abbot and J. Marohasy, “Input selection and optimisation for monthly rainfall forecasting in queensland, australia, using artificial neural networks,” *Atmos. Res.*, vol. 138, pp. 166–178, 2014.
- [61] X. Li, H. R. Maier, and A. C. Zecchin, “Improved PMI-based input variable selection approach for artificial neural network and other data driven environmental and water resource models,” *Environ. Model. Softw.*, vol. 65, pp. 15–29, 2015.
- [62] E. Toth, A. Brath, and A. Montanari, “Comparison of short-term rainfall prediction models for real-time flood forecasting,” *J. Hydrol.*, vol. 239, no. 1–4, pp. 132–147, 2000.
- [63] A. P. Piotrowski and J. J. Napiorkowski, “A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling,” *J. Hydrol.*, vol. 476, pp. 97–111, 2013.
- [64] V. Nourani, M. Komasi, and A. Mano, “A multivariate ANN-wavelet approach for rainfall-runoff modeling,” *Water Resour. Manag.*, vol. 23, no. 14, pp. 2877–2894, 2009.
- [65] F. Mekanik, M. A. Imteaz, S. Gato-Trinidad, and A. Elmahdi, “Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes,” *J. Hydrol.*, vol. 503, pp. 11–21, 2013.
- [66] M. J. Alizadeh, M. R. Kavianpour, O. Kisi, and V. Nourani, “A new approach for simulating and forecasting the rainfall-runoff process within the next two months,” *J. Hydrol.*, vol. 548, pp. 588–597, 2017.
- [67] J. Wu, J. Long, and M. Liu, “Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm,” *Neurocomputing*, 2015.
- [68] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, “Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir,” *J. Hydrol.*, vol. 476, pp. 433–441, 2013.
- [69] R. P. Raymundo and D. M. S. Raymundo, “Artificial Neural Network Model and Multiple Regression Analysis Model in Predicting Rainfall – the Case of Isabela , Philippines,” *J. Arts Sci.*, vol. 5, no. 4, pp. 243–263, 2012.
- [70] G. J. Sawale and S. R. Gupta, “Use of Artificial Neural Network in Data Mining For Weather Forecasting,” *Int. J. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 383–387, 2013.
- [71] N. Acharya *et al.*, “A neurocomputing approach to predict monsoon rainfall in monthly scale using SST anomaly as a predictor,” *Acta Geophys.*, vol. 344, no. 2, pp. 695–707, 2012.
- [72] Suhartono, R. Faulina, D. A. Lusia, B. W. Otok, and H. Kuswanto, “Ensemble Method based on ANFIS-ARIMA for Rainfall Prediction,” in *Statistics in Science, Business, and Engineering (ICSSBE)*, 2012

- International Conference*, 2012, pp. 4–7.
- [73] F. Nhita and Adiwijaya, “A Rainfall Forecasting using Fuzzy System Based on Genetic Algorithm,” *2013 Int. Conf. Inf. Commun. Technol. A*, pp. 111–115, 2013.
  - [74] S. Nurcahyo, F. Nhita, and Adiwijaya, “Rainfall prediction in kemayoran Jakarta using hybrid genetic algorithm (GA) and partially connected feedforward neural network (PCFNN),” *2014 2nd Int. Conf. Inf. Commun. Technol. ICoICT 2014*, pp. 166–171, 2014.
  - [75] S. W. Pratama, “Implementation of Local Regression Smoothing and Fuzzy-Grammatical Evolution on Rainfall Forecasting for Rice Planting Calendar,” vol. 4, no. c, 2016.
  - [76] F. Nhita, D. Saepudin, Adiwijaya, and U. N. Wisesty, “Comparative Study of Moving Average on Rainfall Time Series Data for Rainfall Forecasting Based on Evolving Neural Network Classifier,” *Proc. - 2015 3rd Int. Symp. Comput. Bus. Intell. ISCBI 2015*, pp. 8–12, 2016.
  - [77] K. D. Hartomo and E. Winarko, “ESSPI : Exponential Smoothing Seasonal Planting Index , A New Algorithm For Prediction Rainfall,” vol. 14, no. 6, pp. 1–10, 2016.
  - [78] J. Abbot and J. Marohasy, “Using lagged and forecast climate indices with artificial intelligence to predict monthly rainfall in the brisbane catchment, Queensland, Australia,” *Int. J. Sustain. Dev. Plan.*, vol. 10, no. 1, pp. 29–41, 2015.
  - [79] Y. S. Andrew W. Robertson, Vincent Morona, “Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia,” *Int. J. Climatol.*, vol. 29, no. 27 Nopember 2008, pp. 1449–1462, 2009.
  - [80] A. Schepen, Q. J. Wang, and D. Robertson, “Evidence for using lagged climate indices to forecast Australian seasonal rainfall,” *J. Clim.*, vol. 25, no. 4, pp. 1230–1246, 2012.
  - [81] R. Tresnawati and K. E. Komalasari, “Skenario tenggang waktu sst nino 3.4 terhadap curah hujan untuk meningkatkan akurasi prediksi kalman filter,” *J. Meteorol. dan Geofis.*, vol. 12, no. 3, pp. 243–251, 2011.
  - [82] Y. S. Swarinoto, A. H. Wigena, and D. Statistika, “Statistical Downscaling Suhu Muka Laut Global Untuk Prediksi Total Hujan Bulanan,” *J. Meteorol. dan Geofis.*, vol. 12, pp. 9–19, 2011.
  - [83] A. F. R. H. Harisa Bilhaqqi Qalbi, “Future rainfall variability in Indonesia under different ENSO and IOD composites based on decadal predictions of CMIP5 datasets,” *J. Phys. Conf. Ser.*, vol. 755, p. 11001, 2017.
  - [84] G. B. Humphrey, S. Galelli, a Castelletti, H. R. Maier, G. C. Dandy, and M. S. Gibbs, “A new evaluation framework for input variable selection algorithms used in environmental modelling,” *Int. Environ. Model. Softw. Soc.*, vol. 62, pp. 33–51, 2014.
  - [85] P. Cortez, “Sensitivity Analysis for Time Lag Selection to Forecast Seasonal Time Series using Neural Networks and Support Vector Machines,” 1986.
  - [86] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, “Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles,” *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 9, pp. 2522–2535, 2015.
  - [87] M. S. Park and T. Heo, “Seasonal Spatial-temporal Model for Rainfall Data

- of South Korea,” *J. Appl. Sci. Res.*, vol. 5, no. 5, pp. 565–572, 2009.
- [88] R. H. Kripalani and A. Kulkarni, “Rainfall variability over South-east Asia—connections with Indian monsoon and ENSO extremes: new perspectives,” *Int. J. Climatol.*, vol. 17, no. 11, pp. 1155–1168, 1997.
  - [89] J. Sulaiman, H. Darwis, and H. Hirose, “Forecasting monthly maximum 5-day precipitation using artificial neural networks with initial lags,” *Proc. - 6th Int. Symp. Comput. Intell. Des. Isc. 2013*, vol. 2, pp. 3–7, 2013.
  - [90] H. D. Tran, N. Muttill, and B. J. C. Perera, “Selection of significant input variables for time series forecasting,” *Environ. Model. Softw.*, vol. 64, pp. 156–163, 2015.
  - [91] N. Yuan, Z. Fu, H. Zhang, L. Piao, E. Xoplaki, and J. Luterbacher, “Detrended Partial-Cross-Correlation Analysis: A New Method for Analyzing,” *Sci. Rep.*, pp. 1–7, 2015.
  - [92] X. Qian, Y. Liu, Z. Jiang, B. Podobnik, W. Zhou, and H. E. Stanley, “Detrended partial cross-correlation analysis of two nonstationary time series influenced by common external forces,” *Phys. Rev.*, vol. 62816, pp. 6–11, 2015.
  - [93] R. Rak, S. Drozd, J. Kwapien, and P. Oswiecimka, “Detrended cross-correlations between returns, volatility, trading activity, and volume traded for the stock market companies,” *EPL (Europhysics Lett.)*, vol. 48001, pp. 1–6, 2015.
  - [94] S. Y. Jiang and L. X. Wang, “Efficient feature selection based on correlation measure between continuous and discrete features,” *Inf. Process. Lett.*, vol. 116, no. 2, pp. 203–215, 2016.
  - [95] B. D. Fulcher and N. S. Jones, “Highly comparative feature-based time-series classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 3026–3037, 2014.
  - [96] H. Yoon, K. Yang, and C. Shahabi, “Feature subset selection and feature ranking for multivariate time series,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 9, pp. 1186–1198, 2005.
  - [97] M. Z. Hossain, M. M. Kabir, and M. Shahjahan, “A robust feature selection system with Colin’s CCA network,” *Neurocomputing*, vol. 173, pp. 855–863, 2016.
  - [98] L. Yu and H. Liu, “Efficient Feature Selection via Analysis of Relevance and Redundancy,” *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, 2004.
  - [99] V. Bolón-Canedo, N. Sánchez-Marono, and A. Alonso-Betanzos, “A review of feature selection methods on synthetic data,” *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, 2013.
  - [100] A. Jovic and F. Jovic, “Classification of cardiac arrhythmias based on alphabet entropy of heart rate variability time series,” *Biomedical Signal Processing and Control*, vol. 31, pp. 217–230, 2017.
  - [101] M. Han and X. Liu, “Feature selection techniques with class separability for multivariate time series,” *Neurocomputing*, vol. 110, pp. 29–34, 2013.
  - [102] L. I. U. Tianhong, W. E. I. Haikun, Z. Kanjian, and G. U. O. Weili, “Mutual information based feature selection for multivariate time series forecasting,” in *Proceedings of the 35th Chinese Control Conference*, 2016, pp. 7110–7114.

- [103] Z. Wang, M. Li, and J. Li, "A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure," *Inf. Sci. (Ny)*, vol. 307, pp. 73–88, 2015.
- [104] H. Li, "Accurate and efficient classification based on common principal components analysis for multivariate time series," *Neurocomputing*, vol. 171, pp. 744–753, 2015.
- [105] R. Taormina and K. W. Chau, "Data-driven input variable selection for rainfall-runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines," *J. Hydrol.*, vol. 529, pp. 1617–1632, 2015.
- [106] N. Clements, S. Sarkar, and W. Wei, "Multiplicative Spatio-Temporal Models," *J. Int. Energy Policy*, vol. 3, no. 1, pp. 1–14, 2014.
- [107] Edwin Aldrian and Y. S. Djamil, "Spatio-temporal climatic change of rainfall in East Java Indonesia," *Int. J. Climatol.*, vol. 28, no. 12 June 2007, pp. 435–448, 2008.
- [108] V. Nourani and S. Mousavi, "Spatiotemporal groundwater level modeling using hybrid artificial intelligence-meshless method," *J. Hydrol.*, vol. 536, pp. 10–25, 2016.
- [109] V. Nourani, M. T. Alami, and F. D. Vousoughi, "Self-organizing map clustering technique for ANN-based spatiotemporal modeling of groundwater quality parameters," *J. Hydroinformatics*, p. jh2015143, 2015.
- [110] A. Abraham, D. Steinberg, and N. S. Philip, "Rainfall Forecasting Using Soft Computing Models and Multivariate Adaptive Regression Splines," *IEE SMC Trans.*, vol. 1, pp. 1–6, 2001.
- [111] Z. Beheshti, M. Firouzi, S. M. Shamsuddin, M. Zibarzani, and Z. Yusop, "A new rainfall forecasting model using the CAPSO algorithm and an artificial neural network," *Neural Comput. Appl.*, vol. 27, no. 8, pp. 2551–2565, 2016.
- [112] S. Makridakis and M. Hibon, "The M3-Competition: results, conclusions and implications," *Int. J. Forecast.*, vol. 16, pp. 451–476, 2000.
- [113] J. A. Informasi and D. Redaksi, "Jangan Abaikan Informasi Cuaca," *Info BMKG*, 2013.
- [114] K. B. T. and H. L. M. Solomon, S., D.M.Manning, Z.Chen, Marquis, "Climate Change 2007: The Physical Science Basic. Contribution of working group I to the fourth Assesment report of the intergovernmental panel on Climate Change," *Cambridge University Press, Cambridge, United Kingdom and New York, NY,USA.*, no. October, pp. 1–113, 2010.
- [115] Pusat Database BMKG, "dataonline.bmkg.go.id," 2015.
- [116] Giarno, Z. L. Dupe, and M. A. Mustofa, "Kajian awal musim hujan dan awal musim kemarau di indonesia," *J. Mkg*, vol. 13, no. 1, pp. 1–8, 2012.
- [117] B. M. K. dan Geofisika, *Prakiraan Musim Kemarau 2015 di Indonesia*. Jakarta, 2015.
- [118] D. Syaifullah, "Kajian Sea Surface Temperature (SST), Southern Oscillation Index (SOI) dan Dipole Mode Pada Kegiatan Penerapan Teknologi Modifikasi Cuaca," *J. Sains Teknol. Modif. Cuaca*, vol. 11, no. 1, pp. 1–8, 2010.
- [119] K. Vasimalla, "A Survey on Time Series Data Mining," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 5, pp. 170–179, 2014.



- [120] A. Saikhu, A. Z. Arifin, and C. Fatichah, "Rainfall Forecasting by Using Autoregressive Integrated Moving Average, Single Input and Multi Input Transfer Function," in *ICTS*, 2017.
- [121] O. E. Ayinde, M. Muchie, and G. B. Olatunji, "Effect of Climate Change on Agricultural Productivity in Nigeria : A Co-integration Model Approach," *J. Hum. Ecol.*, vol. 35, no. 3, pp. 189–194, 2011.
- [122] A. Baker and C. Brunsdon, "Non-linearities in drip water hydrology: an example from Stump Cross Caverns, Yorkshire," *J. Hydrol.*, vol. 277, no. 3–4, pp. 151–163, 2003.
- [123] R. May, G. Dandy, and H. Maier, "Review of Input Variable Selection Methods for Artificial Neural Networks."
- [124] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, no. 3, pp. 1157–1182, 2003.
- [125] C. N. Babu and B. E. Reddy, "A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data," *Appl. Soft Comput.*, vol. 23, pp. 27–38, 2014.
- [126] G. P. Reganata, "Inflow and outflow forecasting of currency using multi-input transfer function," *MA.53 Proc. IConSSE FSM SWCU*, pp. 53–65, 2015.
- [127] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [128] J. O. F. Climate and O. Agricultural, "GiST : A Stochastic Model for Generating Spatially and Temporally Correlated Daily Rainfall Data," pp. 5990–6008, 2010.
- [129] R. Srikanthan and T. A. McMahon, "Stochastic generation of annual, monthly and daily climate data: A review," *Hydrol. Earth Syst. Sci.*, vol. 5, no. 4, pp. 653–670, 2001.
- [130] J. G. De Gooijer and R. J. Hyndman, "25 Years of Time Series Forecasting," *Int. J. Forecast.*, vol. 22, no. 3, pp. 443–473, 2006.
- [131] L. Wang, Z. Wang, and S. Liu, "An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm," *Expert Syst. Appl.*, vol. 43, pp. 237–249, 2016.
- [132] N. Challita, M. Khalil, and P. Beausery, "New technique for feature Selection: combination between Elastic Net and Relief Nicole," in *Technological Advances in Electrical, Electronics and Computer Engineering (TAECE)*, 2015, pp. 3–8.
- [133] A. Saikhu, A. Z. Arifin, and C. Fatichah, "Correlation and Symmetrical Uncertainty based Feature Selection for Multivariate Time Series Classification," 2017.
- [134] P. K. Sahu, A. Padhi, G. R. Patil, G. Mahesh, and A. K. Sarkar, "Spatial Temporal Analysis of Major Seaport Freight Flows in India," in *Transportation Research Board 96th Annual Meeting and Possible*, 2017, pp. 1–18.
- [135] D. Palchak, "An Artificial Neural Network in Short-Term Electrical Load Forecasting of a University Campus: A Case Study," *J. Energy Resour. Technol.*, vol. 135, no. 3, p. 32001, 2013.
- [136] T. Cao, E. P. Lim, Z. H. Zhou, T. B. Ho, D. Cheung, and H. Motoda,

- “Advances in knowledge discovery and data mining: 19th pacific-asia conference, PAKDD 2015 Ho Chi Minh City, Vietnam, May 19-22, 2015 proceedings, part I,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9077, pp. 199–211, 2015.
- [137] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [138] A. Saikhu, A. Z. Arifin, and C. Fatichah, “Correlation and Symmetrical Uncertainty based Feature Selection for Multivariate Time Series Classification,” *Draft J. IJASEIT*, 2017.
- [139] D. Gunawan and U. A. Linarka, “Penentuan Prediktor Untuk Prediksi Curah Hujan Bulanan,” *J. Meteorol. dan Geofis.*, no. 2, pp. 93–102.