

IDENTIFIKASI STRUKTUR DEPENDENSI DENGAN COPULA (APLIKASI PADA DATA KLIMATOLOGI)

Mutiah Salamah¹ dan Heri Kuswanto²

^{1,2}Jurusan Statistika, Institut Teknologi Sepuluh Nopember (ITS) Surabaya
e-mail: ¹mutiah_s@statistika.its.ac.id, ²heri_k@statistika.its.ac.id

ABSTRACT

Paper ini mendiskusikan peranan copula dalam memodelkan struktur dependensi antara dua variabel iklim, yaitu kecepatan angin dan rata-rata tekanan udara permukaan laut. Kedua variabel iklim tersebut mempunyai distribusi marginal yang berbeda yang mana kecepatan angin tidak berdistribusi normal dengan kecenderungan heavy-tailed, sedangkan tekanan laut berdistribusi normal. Oleh karena itu dependensi keduanya tidak bisa diwakili dengan menggunakan korelasi Pearson. Masalah ini bisa diatasi dengan copula. Hasil analisis menunjukkan bahwa copula-t adalah model terbaik untuk menjelaskan struktur dependensi antara kedua variabel yang dibahas.

Keywords: *copula, struktur dependensi, heavy-tailed*

PENDAHULUAN

Analisis dependensi antara dua atau lebih variabel merupakan subjek penting dalam aplikasi sehari-hari misalnya marketing, percobaan klinis, ekonomi dan bisnis. Analisis standar biasanya ditampilkan dalam bentuk satu nilai tingkat korelasi, yang kemudian memberikan petunjuk kepada praktisi mengenai kebijakan yang harus diambil. Karena kebijakan yang salah akan mengakibatkan kerugian yang signifikan, maka analisis harus dilakukan secara benar dan teliti.

Paper ini menyajikan suatu metode yang mendapatkan banyak perhatian dalam beberapa tahun terakhir dalam ilmu statistika yaitu pemodelan copula. Ini merupakan suatu metode statistika yang diperuntukkan untuk memodelkan struktur dependensi antara dua atau lebih variabel, dan mempunyai kemampuan untuk mengeksplorasi lebih banyak informasi tentang dependensi daripada korelasi Pearson, Kendall atau Spearman. Dalam prakteknya, seringkali praktisi mengabaikan atau bahkan tidak mengetahui distribusi marginal dari variabel yang dianalisis. Korelasi Pearson seringkali menjadi pilihan paling mudah dan sederhana untuk mengukur dependensinya.

Peranan copula menjadi penting ketika satu atau kedua variabel mempunyai distribusi marginal yang tidak normal atau mempunyai tail dependensi. Sebagaimana yang kita ketahui, korelasi Pearson diperkenalkan dengan mengasumsikan bahwa variabelnya berdistribusi normal. Banyak distribusi bivariat atau multivariat yang dikembangkan sebagai alternatif untuk mengatasi kasus ketidak-

normalan pada distribusi marginal seperti bivariate gamma (Moran (1969), Nadarajah dan Gupta (2006)), Farlie-Gumbel-Morgenster (Conway (1979), bivariate exponential distribution (Gumbel (1960)), dan sebagainya. Namun, distribusi-distribusi tersebut terbatas pada marginal yang sama, dan mempunyai struktur yang kompleks pada fungsi densitas probabilitasnya (termasuk struktur dependensinya) yang tentunya tidak disukai dalam praktek.

Copula datang dengan ke-fleksibilitasnya, dimana distribusi marginal dari variabel bisa berbeda atau bahkan tidak diketahui. Copula telah dapat diaplikasikan dengan baik pada banyak bidang seperti hidrologi (Favre et. al. (2004), Salvadori dan De Michele (2007), Genest et. al. (2007)), finance dan asuransi (Cerubini et. al. (2004), Denuit et. al. (2005), McNeil et.al. (2005)) dan masih banyak lagi. Akhir-akhir ini, copula telah dikembangkan pada bidang ekonometrik dan time series seperti pada Patton (2002;2009) dan lainnya.

Paper ini membahas aplikasi copula dalam memodelkan struktur dependensi antara dua variabel iklim yaitu kecepatan angin dan tekanan udara permukaan laut. Aplikasi copula dalam bidang meteorologi masih sangat sedikit dan tergolong masih baru. Beberapa diantaranya adalah De Michele dan Salvadori (2003), Vannitsem (2007), Vrac et. al. (2005) and Schölzel dan Friedrich (2008).

KONSEP DASAR COPULA

Bagian ini menjelaskan secara singkat tentang konsep dasar copula. Deskripsi yang diberikan terbatas pada dua variabel random atau kasus bivariat. Teori yang digambarkan disini tentu saja dapat dikembangkan untuk kasus multivariat dengan dua atau lebih variabel random.

Misalkan kita punya 2 vektor random X dengan kumulatif distribusi marginal F_{X1} dan F_{X2} , maka berdasarkan teorema Sklar (1959), distribusi bersama dari vektor random dapat dituliskan sebagai fungsi dari marginal distribusinya sebagai berikut

$$F_X(x) = C_X(F_{X1}(x1), F_{X2}(x2))$$

dimana $C_X: [0,1] * [0,1] \rightarrow [0,1]$ adalah fungsi distribusi bersama dari variabel random yang ditransformasi $U_j = F_{Xj}(Xj)$ untuk $j = 1,2$. Transformasi ini menghasilkan distribusi marginal uniform U_j . C_X adalah fungsi copula dengan densitasnya adalah c_X .

Teorema Sklar dia atas mempunyai dua implikasi penting, yaitu probabilitas desitas bersama dapat di ekspresikan sebagai perkalian antara desitas marginal dan desitas copula sebagai berikut:

$$f_X(x) = f_{X1}(x1)f_{X2}(x2) \cdot c_X(u_1, u_2)$$

Berbagai fungsi copula telah dikembangkan sejauh ini. Namun, kita batasi diskusi dalam paper ini pada tiga copula yang terkait dengan kasus yang kita bahas. Mengenai teori copula lebih detail bisa ditemukan pada Genest dan Favre (2007), Nielsen (1999) atau Joe (1997).

a. Copula Gumbel dan Clayton

Copula Gumbel dan Clayton termasuk dalam copula Archimedian, yang memungkinkan mempunyai struktur dependensi yang lebih luas. Dengan copula Archimedian, beberapa copula dapat di bangiktan dengan fungsi generator. Dalam bentuk umum, copula Archimedian mempunyai bentuk sebagai berikut:

$$C_X(u_1, u_2) = \phi^{-1}(\phi(u_1), \phi(u_2))$$

Dimana $\phi(u)$ adalah fungsi generator. $\phi(u)$ adalah fungsi tidak turun yang memetakan $[0,1]$ ke dalam sehingga $\phi(0) = \infty$ and $\phi(1) = 0$. Copula Gumbel mempunya fungsi generator sebagai berikut:

$$\phi(u) = (-\log(u))^\theta, \theta \geq 1$$

Sedangkan copula Clayton

$$\phi(u) = \frac{u^{-\theta} - 1}{\theta}, \theta \geq 0.$$

Kedua copula di atas mempunyai perilaku tail yang berbeda. Copula Clayton mempunyai tail dependensi bawah sedangkan copula Gumbel mempunyai tail dependensi atas.

b. Copula-t

Copula-t termasuk dalam copula elips sebagaimana copula Gaussian (copula dengan marginal normal). Namun tidak seperti copula normal yang simetri, copula-t mempunyai potensi untuk membangkitkan nilai ekstrim karena t adalah distribusi yang skew. Copula-t didefinisikan sebagai berikut:

$$C_X(u_1, u_2) = F_{t(v,\Sigma)}(F_{t(v)}^{-1}(u_1), F_{t(v)}^{-1}(u_2))$$

Dimana F_t mendefinisikan CDF dari distribusi t dengan derajat bebas v .

Copula digunakan secara bersamaan dengan korelasi tau kendall dan spearman correlation. Ini karena kedua korelasi tersebut diurunkan dari rank, yang sesuai dengan teori copula. Copula sendiri tidak berubah dengan adanya transformasi.

DESKRIPSI DATA DAN PENTINGNYA MEMPELAJARI VARIABEL YANG DIBAHAS

Dalam paper ini kita tertarik untuk mengaplikasikan metode copula pada kasus klimatologi. Kita tertarik untuk mempelajari struktur dependensi antara kecepatan angin dengan rata-rata tekanan udara permukaan laut (mslp). Data diperoleh dari stasiun surabaya / juanda. Kita menganalisa data harian selama 9 tahun antara tahun 2000 sampai 2009. Ada beberapa pengamatan yang missing dan mereka dihilangkan dari analisis. Jumlah data keseluruhan sebanyak 1400 data berpasangan. Kecepatan angin diukur dengan km/jam sedangkan mslp diukur dalam mbar.

Sekarang kita mendiskusikan pentingnya menganalisis kedua variabel tersebut. Kecepatan angin (bersamaan dengan arah angin) merupakan variabel penting dalam beberapa disiplin ilmu seperti meteorologi, penerbangan, kelautan, konstruksi dan lain sebagainya. Kecepatan angin yang tinggi mempunyai hubungan dengan petir yang hebat serta angin puting beliung yang bisa menyebabkan kerusakan fatal dan kerugian dalam hidup. Ini juga merupakan input utama dari pembangkit tenaga angin. Berbagai penelitian dalam klimatologi dan meteorologi ditujukan untuk memprediksi kecepatan angin. Mempelajari mslp sangat penting terutama untuk bidang yang sensitif terhadap cuaca, seperti pilot, petani, pelaut, dan sebagainya. Kecepatan angin dan

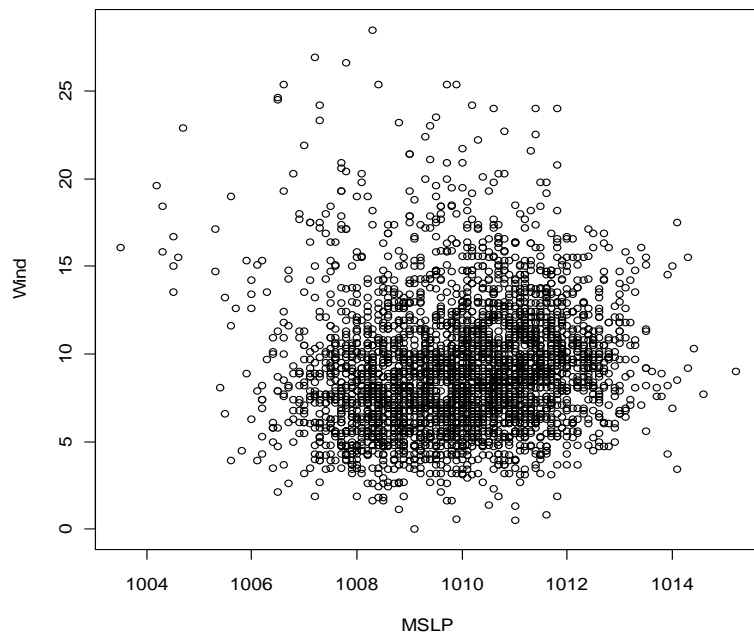
mslp merupakan suatu indikator penting dalam perubahan iklim (harrison dan larkin (1997).

Mempelajari struktur dependensi antara kedua variabel tersebut juga subjek yang penting. Mengembangkan sistem peramalan (model) untuk kecepatan angin membutuhkan mslp sebagai salah satu input. Selanjutnya, kedua variabel ini juga merupakan input untuk peramalan variabel iklim yang lain seperti curah hujan, presipitasi dan petir sebagaimana diuraikan dalam qian et al. (2000), murphree and van den dool (1988), kutiel (2004). Dengan mempertimbangkan struktur dependensi antar variabel tersebut akan dapat memperoleh peramalan yang lebih reliabel. Sebaliknya,

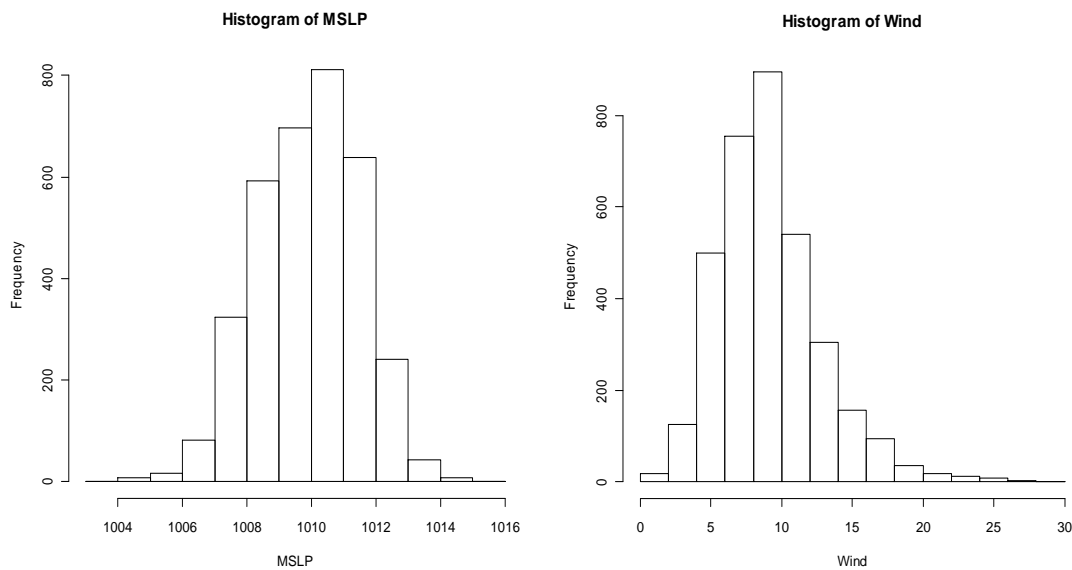
mengabaikannya akan menghasilkan analisis yang salah.

APLIKASI DAN PEMBAHASAN

Gambar berikut menampilkan scatter plot antara variabel yang dibahas dalam skala aslinya. Kita tidak melihat adanya trend atau tendensi tertentu dari plot. Untuk menyimpulkan apakah kedua variabel berkorelasi atau tidak sangat sulit jika kita hanya menggunakan scatter plot. Salah satu informasi penting yang bisa dibaca dari gambar 1 adalah, bahwa plot-plotnya terkonsentrasi pada interval tertentu, yang mengindikasikan dependensi antara kedua variabel tersebut.



Gambar 1. Scatter plot antara kecepatan angin dan MSLP dalam skala asli



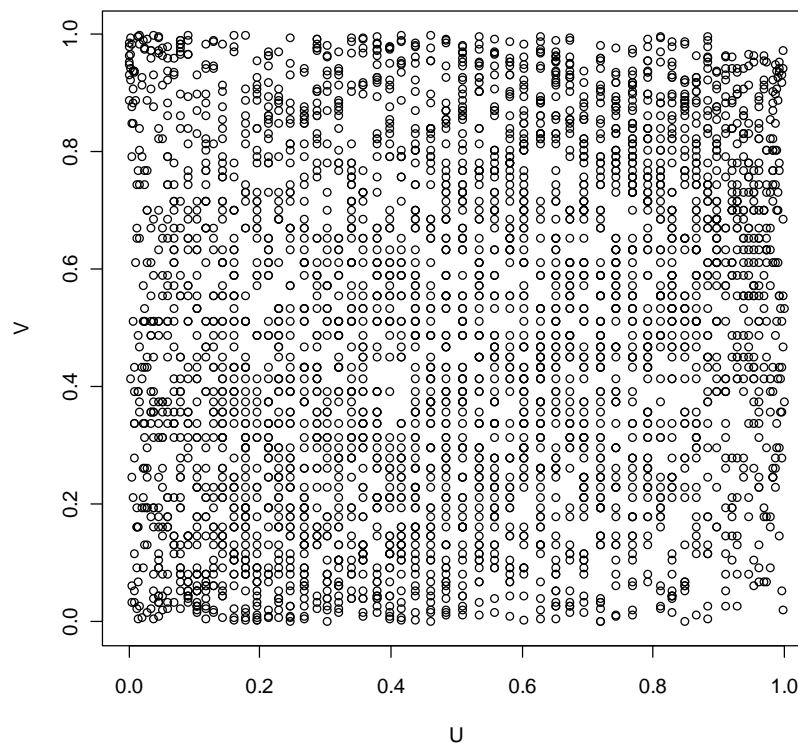
Gambar 2. Histogram (distribusi marginal) dari masing-masing variabel

Meskipun pengetahuan tentang distribusi marginal tidak disyaratkan dalam pemodelan copula, informasi tentang distribusi marginal bermanfaat untuk menjamin bahwa copula memang diperlukan. Jika kedua variabel mempunyai marginal normal, maka korelasi Pearson dapat diaplikasikan dengan mudah. Kita dapat menduga distribusi marginal secara empiris dengan menggunakan histogram. (Gambar 2).

Histogram sebelah kiri menggambarkan histogram dari MSLP dan sebelah kanan mewakili histogram dari kecepatan angin. Dari gambar, terlihat bahwa MSLP sepertinya simetri dan berdistribusi normal, sedangkan kecepatan angin tidak simetri dan mempunyai tail yang panjang.

Kenyataan bahwa salah satu variabel tidak berdistribusi normal menyebabkan struktur dependensi dari kedua variabel itu tidak dapat diwakili oleh korelasi Pearson. Sebagaimana yang dijelaskan sebelumnya, copula mempunyai kemampuan untuk mendeskripsikan struktur dependensi antara variabel dengan marginal yang berbeda dan memodelkan dependensi tail-nya.

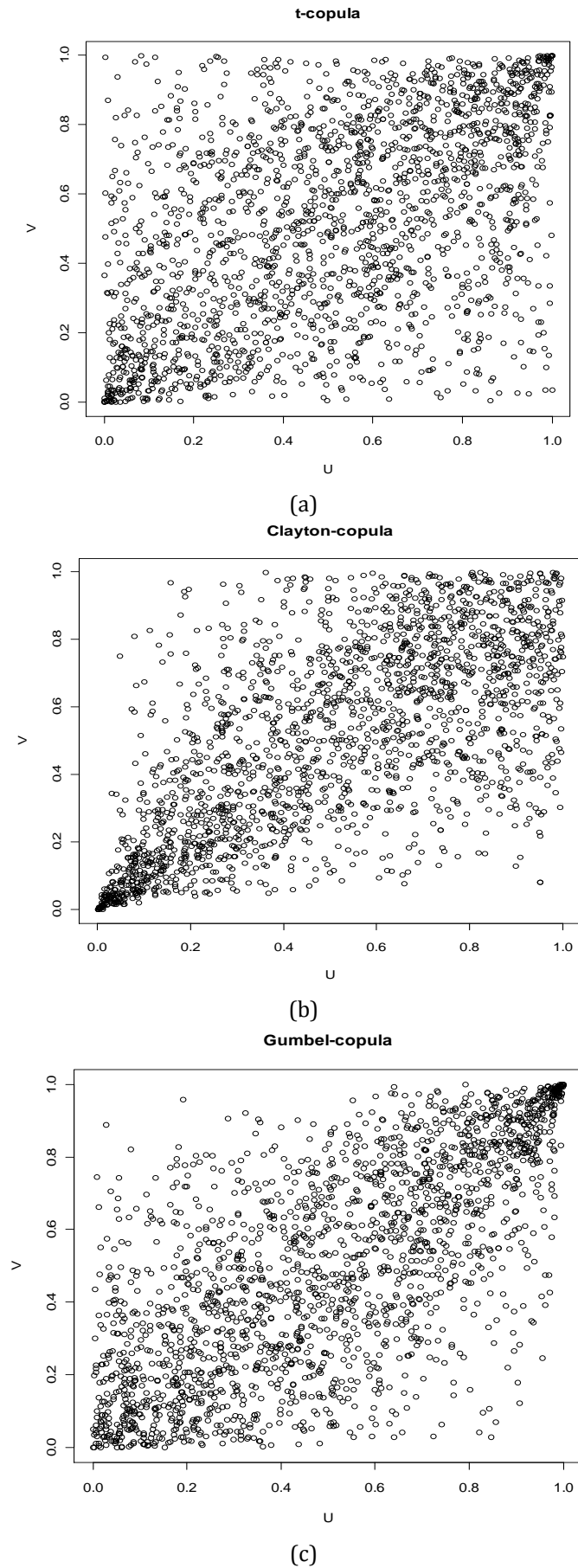
Langkah pertama dalam analisis dengan copula adalah mentransformasi variable ke dalam distribusi marginal yang uniform. Scatter plot pada gambar 1 kemudian di transformasi ke domain $[0,1]$, seperti yang bisa dilihat pada gambar berikut:



Gambar 3. Scatter plot antara kecepatan angin dan MSLP dalam skala transformasi (uniform $[0,1]$)

Plot dalam gambar mengkonfirmasi bahwa kedua variabel saling berhubungan, meskipun derajat hubungannya sangat rendah. Ini tampak jelas dari plot-plot yang terpusat pada pojok, namun tidak jelas apakah pojok bawah atau atas. Bagaimanapun juga, ini suatu indikasi dari dependensi tail. Seperti yang didiskusikan sebelumnya, paling tidak ada 3 copula yang mempunyai karakteristik seperti ini yaitu copula-t, Clayton dan Gumbel.

Gambar 4 mengilustrasikan suatu kasus ideal dengan plot yang dibangkitkan dari 5000 pengamatan. Dari plot tersebut, nampak jelas bahwa copula mempunyai karakteristik sendiri terkait dengan dependensi tail. Untuk meyakinkan copula mana yang cocok dengan kasus kita, seleksi model perlu dilakukan.



Gambar 4. Sampel plot untuk (a)copula-t, (b) Clayton, dan (c) Gumbel

Sebelum kita melangkah pada penaksiran model copula, perlu ditampilkan koefisien korelasi hasil estimasi dengan 3 metode yaitu Pearson, kendall and Spearman.

Tabel 1. Koefisien korelasi

	Pearson	Kendall	Spearman
Correlation	0.09399	0.1228	0.1793

Ketiga ukuran tersebut memberikan nilai yang sangat kecil. Khususnya korelasi Pearson, nilai korelasinya paling kecil diantara yang lain. Menggunakan ukuran korelasi ini akan mengakibatkan kesimpulan yang bias, yaitu kedua variabel yang dianalisis tidak berhubungan.

Dalam praktek, kondisi ini akan menyarankan untuk mengabaikan korelasi antara keduanya (jika digunakan sebagai model input dalam peramalan klimatologi) atau mengeluarkan MSLP sebagai variabel independen dari kecepatan angin (jika kita ingin meramalkan kecepatan angin). Perlu dicatat bahwa korelasi yang lain dihitung dari data yang sudah ditransformasi, dan memberikan nilai yang lebih besar secara signifikan daripada korelasi Pearson.

Untuk meyakinkan apakah nilai korelasi tersebut signifikan atau tidak, perlu dilakukan pengujian hipotesa. Hasil pengujian menunjukkan bahwa kedua variabel tersebut berkorelasi secara signifikan (dengan 95% confidence interval) dengan $p\text{-value} < 2.2e-16$ untuk Kendall dan Spearman. Oleh karena itu, kita harus mempertimbangkan korelasi tersebut. Yang belum teridentifikasi adalah tentang peranan copula dan copula mana yang mewakili struktur dependensi dengan baik. Berikut hasil estimasi parameter copula untuk tiga macam copula:

Tabel 2. Estimasi model copula

Copula	Estimasi	Z	log likelihood
t	0.1585	9.514	41.76
Gumbel	1.0795	93.56	21.48
Clayton	0.1644	8.4185	29.754

Statistik Z menyarankan bahwa parameter dari ketiga copula tersebut signifikan. Estimasinya ditampilkan dengan tau Kendall. Model terbaik akan diseleksi berdasarkan nilai log-likelihood. Karena nilai ini mewakili kemungkinan maksimum bahwa model yang diestimasi akan cocok dengan model sebenarnya, maka semakin besar log likelihood akan semakin baik modelnya. Dari tabel, copula-t memberikan nilai paling tinggi. Oleh karena itu, kita menyimpulkan bahwa struktur dependensi antara kecepatan angin dan MSLP lebih bagus

diwakili oleh copula-t dengan parameter 0.1585 bersesuaian dengan 0.1228. Hasil ini menyarankan bahwa kita harus memperhatikan nilai ekstrim (atas atau bawah). Kedua variabel secara signifikan saling bergantung pada kondisi ekstrim rendah atau tinggi. Pengujian kesesuaian model tidak dibahas dalam paper ini.

PENUTUP

Kita telah menunjukkan aplikasi copula untuk memodelkan struktur dependensi antara kecepatan angin dengan MSLP. Korelasi Pearson tidak lagi dapat dipakai karena distribusi marginal dari keduanya berbeda. Copula dapat menangkap dependensi tail antara kedua variabel dan menyarankan copula-t sebagai model terbaik. Hasil yang ditampilkan dalam paper ini adalah kasus khusus per stasiun. Oleh sebab itu, stasiun lain mungkin akan mempunyai struktur dependensi yang berbeda.

DAFTAR PUSTAKA

- [1] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*, Wiley.
- [2] Conway, D.A. (1979). *Farlie-Gumbel-Morgenstern distributions*. In: Encyclopedia of Statistical Sciences, 3, S.Kotz and N.L. Johnson, editors (John Wiley & Sons, New York), 28-31.
- [3] De Michele, C. and Salvadori, G. (2003). A Generalized Pareto Intensity Duration Model of Storm Rainfall Exploiting 2-Copulas, *J. Geophys. Res.*, 108(D2); 4067.
- [4] Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for dependent Risks. Measures, Orders and Models*. Wiley, Chichester.
- [5] Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling Dependence with Copulas and Applications to Risk Management, in: *Handbook of Heavy Tailed Distributions in Finance*, edited by: Rachev, S., 329-384.
- [6] Genest, C. and Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask, *J. Hydrol. Eng.*; 12; 347-368.
- [7] Genest, C., Favre, A.-C., Beliveau, J., and Jacques, C. (2007). Metaelliptical Copulas and Their Use in Frequency Analysis of Multivariate Hydrological Data, *Water Resour. Res.*, 43; W09401

- [8] Gumbel, E.J. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association*; 55; 698-707.
- [9] Harrison, D. E. and Larkin, N. K. (1997). Darwin Sea Level Pressure: 1876-1996 : Evidence for Climate Change? *Geophysical research letters*, 24(14); 1779-1782.
- [10] Joe, H. (1997), *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- [11] Kutiel, T. (1991). Recent Spatial and Temporal Variations in Mean Sea Level Pressure over Europe and the Middle East, and Their Influence on the Rainfall Regime in the Galilee, Israel. *Theoretical and Applied Climatology*, 44(3-4).
- [12] McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton.
- [13] Moran, P. A. P. (1969). Statistical Inference with Bivariate Gamma Distributions, *Biometrika*; 54; 385-394.
- [14] Murphree, T., and Van Den Dool, H., (1988). Calculating Tropical Winds from Time Mean Sea Level Pressure Fields. *Journal of the Atmospheric Sciences*, 45; 3269-3282
- [15] Nadarajah, S. and Gupta, A. K. (2006). Some Bivariate Gamma Distributions, *Applied Mathematics Letters*; 19(8); 767-774.
- [16] Patton A. J. (2002). *Applications of Copula Theory in Financial Econometrics*, June Ph.D. dissertation [Department of Economics, University of California, San Diego](#)
- [17] Patton, A. J. (2009) . Copula-Based Models for Financial Time Series, 2009, in [T.G. Andersen](#), [R.A. Davis](#), J.-P. Kreiss and [T. Mikosch](#) (eds.) *Handbook of Financial Time Series*, Springer Verlag
- [18] Nelsen, R.B. (1999), *An Introduction to Copulas*. Lecture Notes in Statistics 139, Springer-Verlag, New York.
- [19] Qian, B., Corte- Real, J. and Xu, H. (2000). Nonseasonal Variability of Monthly Mean Level Pressure and Precipitation Variability over Europe. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*; 25(2);177-181.
- [20] Salvadori , G. and De Michele, C. (2007), On the Use of Copulas in Hydrology: Theory and Practice. *Journal of Hydrologic Engineering*, 12(4); 369-380
- [21] Schölzel, C. and Friedrich, P. (2008). Multivariate non-normally Distributed RandomVariable in Climate Research – Introduction to the Copula Approach. *Nonlinear Process Geophys.*; 25; 761-772.
- [22] Sklar, A. (1959). Fonctions de Repartition an Dimensions et Leurs Marges. *Publ. Inst. Statist. Univ. Paris* 8, 229-231.
- [23] Vannitsem, S. (2007). Statistical Properties of the Temperature Maxima in an Intermediate Order Quasi-Geostrophic model, *Tellus A*; 59; 80-95.
- [24] Vrac, M., Chedin, A., and Diday, E.(2005). Clustering a Global Field of Atmospheric Profiles by Mixture Decomposition of Copulas, *J. Atmos. Ocean. Tech.*; 22; 1445-1459.