ulm university universität uulm

Probabilistic prediction of solar power supply to distribution networks, using global radiation forecasts

Volker Schmidt

Ulm University, Institute of Stochastics

2nd ISM-UUlm Joint Workshop, Oktober 10, 2019

Outline

Outline

Risk in feed-in of solar power

Motivation

- Increase in solar plants $\rightarrow$ voltage violations and overloading problems

Risk in feed-in of solar power

Motivation

- Increase in solar plants → voltage violations and overloading problems
- Solar plants are curtailed → High costs and loss of energy

Risk in feed-in of solar power

Motivation

- Increase in solar plants → voltage violations and overloading problems
- Solar plants are curtailed → High costs and loss of energy
- Data-based predictions might reduce unnecessary curtailment

Risk in feed-in of solar power

Motivation

- Increase in solar plants → voltage violations and overloading problems
- Solar plants are curtailed → High costs and loss of energy
- Data-based predictions might reduce unnecessary curtailment

Data

Timeframe: May, June and July of the years 2015-2017 (11-12 UTC)

Risk in feed-in of solar power

Motivation

- Increase in solar plants → voltage violations and overloading problems
- Solar plants are curtailed → High costs and loss of energy
- Data-based predictions might reduce unnecessary curtailment

Data

Timeframe: May, June and July of the years 2015-2017 (11-12 UTC)

- Global radiation forecasts generated by Deutscher Wetterdienst (DWD)

Risk in feed-in of solar power

Motivation

- Increase in solar plants → voltage violations and overloading problems
- Solar plants are curtailed → High costs and loss of energy
- Data-based predictions might reduce unnecessary curtailment

Data

Timeframe: May, June and July of the years 2015-2017 (11-12 UTC)

- Global radiation forecasts generated by Deutscher Wetterdienst (DWD)
- Solar power supply measured by a distribution network operator in Northern Bavaria (MDN)

Risk in feed-in of solar power

Motivation

- Increase in solar plants → voltage violations and overloading problems
- Solar plants are curtailed → High costs and loss of energy
- Data-based predictions might reduce unnecessary curtailment

Data

Timeframe: May, June and July of the years 2015-2017 (11-12 UTC)

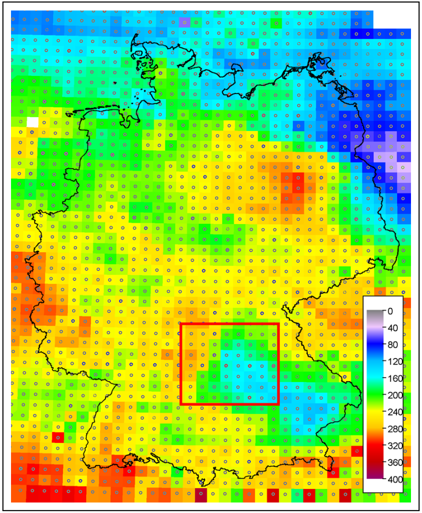
- Global radiation forecasts generated by Deutscher Wetterdienst (DWD)
- Solar power supply measured by a distribution network operator in Northern Bavaria (MDN)

Goal

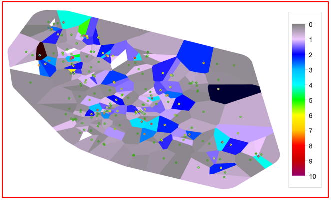
Predict the risk of solar power supply exceeding critical thresholds

Outline

Visualization of data

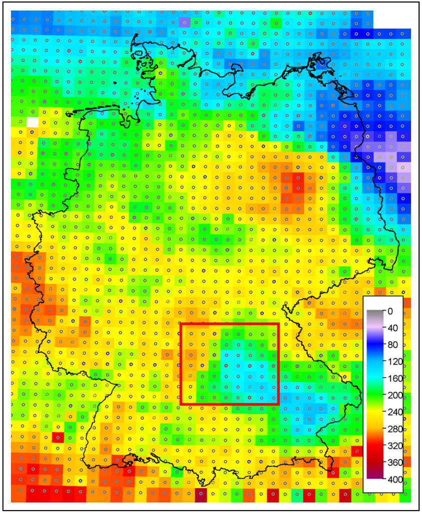


Global radiation forecast (in J/cm^2) for July 07, 2017 11-12 UTC

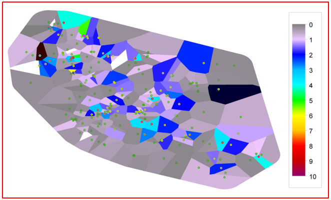


Measured solar power supply (in *MW*) for July 07, 2017 11-12 UTC

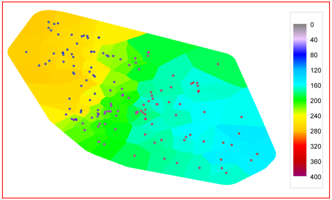
Visualization of data



Global radiation forecast (in J/cm^2)
for July 07, 2017 11-12 UTC

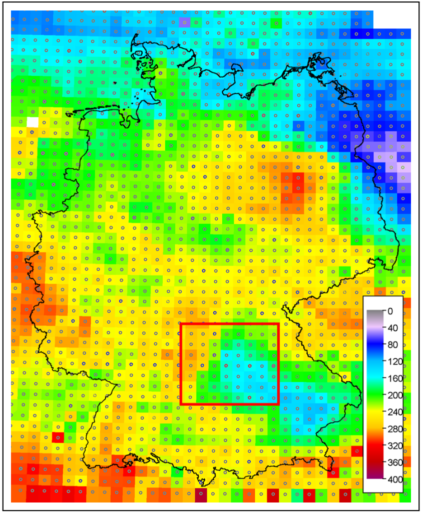


Measured solar power supply (in *MW*)
for July 07, 2017 11-12 UTC

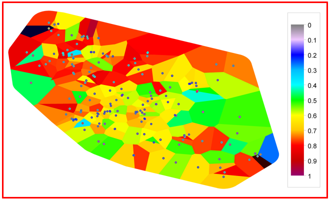


Interpolated global radiation forecast
for July 07, 2017 11-12 UTC

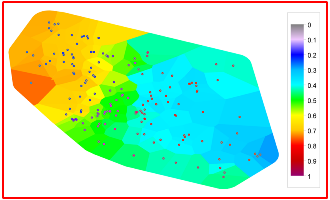
Visualization of data



Global radiation forecast (in J/cm^2) for July 07, 2017 11-12 UTC



Normalized solar power supply for July 07, 2017 11-12 UTC



Normalized global radiation forecast for July 07, 2017 11-12 UTC

Outline

Copula models

Random variables

R: (Normalized) global radiation forecast
S: (Normalized) solar power supply

Copula models

Random variables

R: (Normalized) global radiation forecast
S: (Normalized) solar power supply

Goals

For a predefined threshold v and feed-in points $p_1, \ldots, p_n$ compute the conditional probabilities

- $P(S_1 \geq v \mid R_1 = r(p_1, t))$
- $P(S_1 + \ldots + S_n \geq v \mid R_1 = r(p_1, t), \ldots, R_n = r(p_n, t))$

given global radiation forecasts $r(p_1, t), \ldots, r(p_n, t)$ and forecast time t

Copula models

Random variables

R: (Normalized) global radiation forecast
S: (Normalized) solar power supply

Goals

For a predefined threshold v and feed-in points $p_1, \ldots, p_n$ compute the conditional probabilities

- $P(S_1 \geq v \mid R_1 = r(p_1, t))$
- $P(S_1 + \ldots + S_n \geq v \mid R_1 = r(p_1, t), \ldots, R_n = r(p_n, t))$

given global radiation forecasts $r(p_1, t), \ldots, r(p_n, t)$ and forecast time t

Modeling approach

- Fit univariate marginal distributions
- Fit bivariate and multivariate distributions using bivariate copulas and D-vine copulas

Outline

Copula theory

Bivariate copulas

A bivariate copula is the joint distribution function $C : [0,1] \times [0,1] \to [0,1]$ of a 2-dimensional random vector (U, V) with components U and V uniformly distributed on $[0,1]$

Copula theory

Bivariate copulas

A bivariate copula is the joint distribution function $C : [0,1] \times [0,1] \to [0,1]$ of a 2-dimensional random vector (U, V) with components U and V uniformly distributed on $[0,1]$

Theorem of Sklar

Let (R, S) be a 2-dimensional random vector with joint distribution function $F_{(R,S)} : \mathbb{R}^2 \to [0,1]$ and marginal distribution functions F_R and F_S. Then, a bivariate copula function $C : [0,1] \times [0,1] \to [0,1]$ exists such that

$$F_{(R,S)}(r,s) = C(F_R(r), F_S(s)) \qquad \text{for all } r, s \in \mathbb{R}$$

Copula theory

Bivariate copulas

A bivariate copula is the joint distribution function $C : [0,1] \times [0,1] \to [0,1]$ of a 2-dimensional random vector (U, V) with components U and V uniformly distributed on $[0,1]$

Theorem of Sklar

Let (R, S) be a 2-dimensional random vector with joint distribution function $F_{(R,S)} : \mathbb{R}^2 \to [0,1]$ and marginal distribution functions F_R and F_S. Then, a bivariate copula function $C : [0,1] \times [0,1] \to [0,1]$ exists such that

$$F_{(R,S)}(r,s) = C(F_R(r), F_S(s)) \qquad \text{for all } r, s \in \mathbb{R}$$

Differential form of Sklar's theorem

For the density functions $f_{(R,S)}$, f_R, f_S and c it holds that

$$f_{(R,S)}(r,s) = f_R(r) \cdot f_S(s) \cdot c(F_R(r), F_S(s)) \qquad \text{for all } r, s \in \mathbb{R}$$

Outline

Definition of Archimedean copulas

Archimedean generator

A function $g : [0, 1] \to [0, \infty]$ is called Archimedean generator if g is continuous, strictly decreasing and solves $g(1) = 0$.

Definition of Archimedean copulas

Archimedean generator

A function $g : [0, 1] \to [0, \infty]$ is called Archimedean generator if g is continuous, strictly decreasing and solves $g(1) = 0$.

Pseudo-inverse

The pseudo-inverse $g^{[-1]}$ of an Archimedean generator g is an extension of the inverse function $g^{(-1)}$ defined as

$$g^{[-1]}(t) = \begin{cases} g^{(-1)}(t), & \text{if } 0 \leq t \leq g(0) \\ 0, & \text{if } g(0) \leq t \leq \infty. \end{cases}$$

Definition of Archimedean copulas

Archimedean generator

A function $g : [0,1] \to [0,\infty]$ is called Archimedean generator if g is continuous, strictly decreasing and solves $g(1) = 0$.

Pseudo-inverse

The pseudo-inverse $g^{[-1]}$ of an Archimedean generator g is an extension of the inverse function $g^{(-1)}$ defined as

$$g^{[-1]}(t) = \begin{cases} g^{(-1)}(t), & \text{if } 0 \leq t \leq g(0) \\ 0, & \text{if } g(0) \leq t \leq \infty. \end{cases}$$

Arichmedian copula

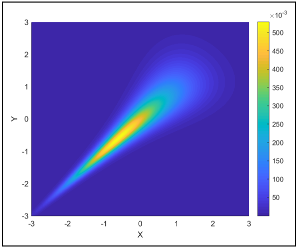
The Arichmedian copula generated by g is given by

$$C(u,v) = g^{[-1]}(g(u) + g(v)) \qquad u, v \in [0,1].$$

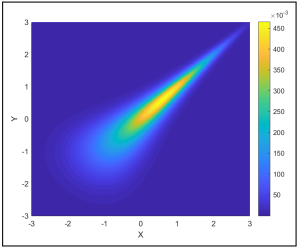
Examples of Archimedean copulas

copula family	Archimedean generator	parameter
Joe	$g(t) = -\log(1-(1-t)^{\theta})$	$\theta \in [1,\infty]$
Frank	$g(t) = (-\log(\frac{\exp(-\theta t-1)}{\exp(-\theta)-1}))^{\theta}$	$\theta \in \mathbb{R}\setminus\{0\}$
Clayton	$g(t) = \frac{1}{\theta}(t^{-\theta}-1)$	$\theta \in [-1,\infty)\setminus\{0\}$
Gumbel	$g(t) = (-\log(t))^{\theta}$	$\theta \in [1,\infty)$

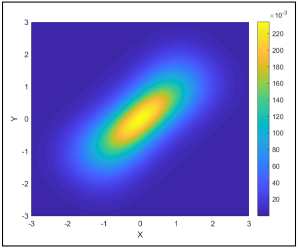
Visualization of the copula types



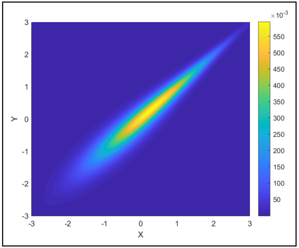
Clayton with parameter $\theta = 5$



Joe with parameter $\theta = 5$



Frank with parameter $\theta = 5$



Gumbel with parameter $\theta = 5$

Outline

Fitting of univariate marginal distributions

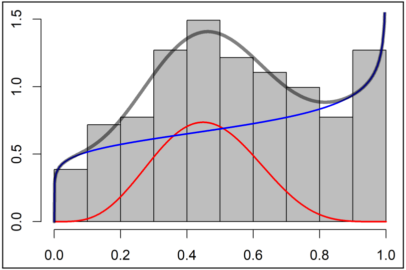
- Fit mixed beta densities $f(x) = qf_1(x) + (1-q)f_2(x)$ with mixture parameter $q \in [0,1]$ and beta densities $f_i : (0,1) \to [0,\infty)$ with $f_i(x) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} x^{a_i-1}(1-x)^{b_i-1}$ and two parameters $a_i, b_i > 0$

Fitting of univariate marginal distributions

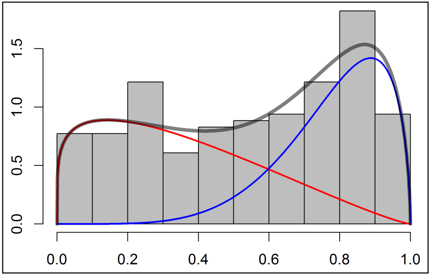
- Fit mixed beta densities $f(x) = qf_1(x) + (1-q)f_2(x)$ with mixture parameter $q \in [0,1]$ and beta densities $f_i : (0,1) \to [0,\infty)$ with $f_i(x) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} x^{a_i-1}(1-x)^{b_i-1}$ and two parameters $a_i, b_i > 0$
- Apply EM algorithm to estimate the parameters of the mixed beta densities

Fitting of univariate marginal distributions

- Fit mixed beta densities $f(x) = qf_1(x) + (1-q)f_2(x)$ with mixture parameter $q \in [0,1]$ and beta densities $f_i : (0,1) \to [0,\infty)$ with $f_i(x) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} x^{a_i-1}(1-x)^{b_i-1}$ and two parameters $a_i, b_i > 0$
- Apply EM algorithm to estimate the parameters of the mixed beta densities



Global radiation forecast data



Solar power supply data

Fitting bivariate copulas

- Estimate the copula parameter θ for each copula type by maximizing the likelihood function with given F_R and F_S

Fitting bivariate copulas

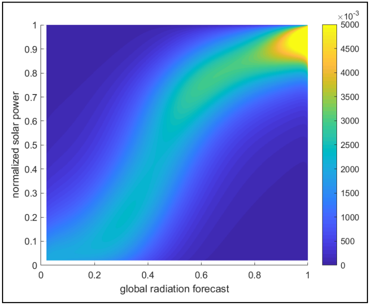
- Estimate the copula parameter θ for each copula type by maximizing the likelihood function with given F_R and F_S
- Choose the copula C by maximizing the likelihood function over all copula types

Fitting bivariate copulas

- Estimate the copula parameter θ for each copula type by maximizing the likelihood function with given F_R and F_S
- Choose the copula C by maximizing the likelihood function over all copula types
- The Frank copula gives us the best fit

Fitting bivariate copulas

- Estimate the copula parameter θ for each copula type by maximizing the likelihood function with given F_R and F_S
- Choose the copula C by maximizing the likelihood function over all copula types
- The Frank copula gives us the best fit



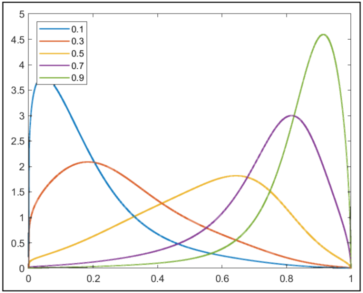
Fitted bivariate joint density

Outline

Computation of conditional probabilities

Computation of conditional probabilities

- Based on Sklar's theorem compute the conditional density
$f_S(s \mid R = r) = \frac{f_{(R,S)}(r,s)}{f_R(r)} = f_S(s) \cdot c(F_R(r), F_S(s))$

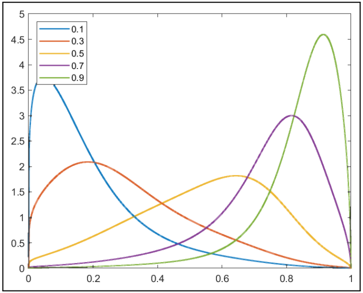


Conditional densities given the global radiation forecasts r for $r = 0.1, 0.3, 0.5, 0.7$ and 0.9

Computation of conditional probabilities

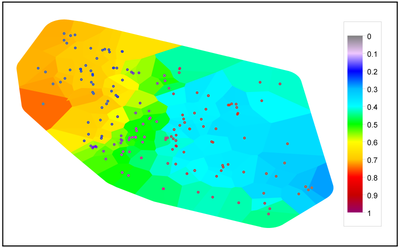
Computation of conditional probabilities

- Based on Sklar's theorem compute the conditional density
  $f_S(s \mid R = r) = \frac{f_{(R,S)}(r,s)}{f_R(r)} = f_S(s) \cdot c(F_R(r), F_S(s))$
- Compute the conditional probabilities
  $P(v, r) = P(S \geq v \mid R = r) = \int_v^1 f_S(s \mid R = r)\mathrm{d}s$

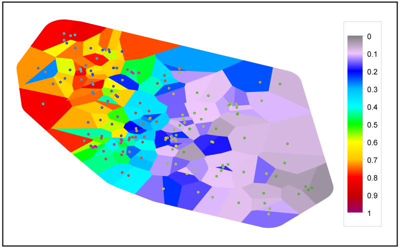


Conditional densities given the global radiation forecasts r for $r = 0.1, 0.3, 0.5, 0.7$ and 0.9

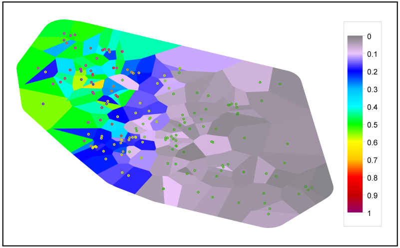
Conditional probabilities for a critical event



Normalized global radiation forecast

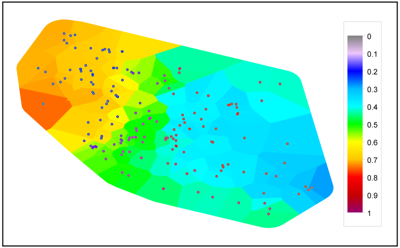


Conditional probabilities for threshold $v = 0.7$

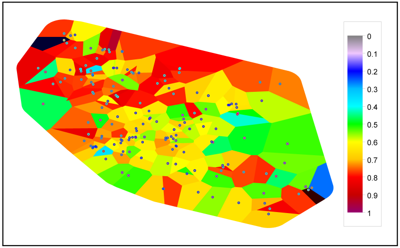


Conditional probabilities for threshold $v = 0.8$

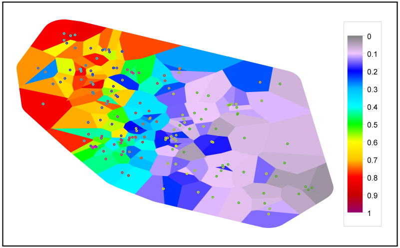
Conditional probabilities for a critical event



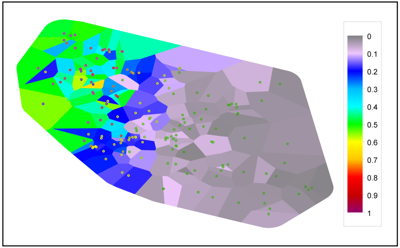
Normalized global radiation forecast



Normalized solar power supply



Conditional probabilities for threshold $v = 0.7$



Conditional probabilities for threshold $v = 0.8$

Outline

Vine copulas

Multivariate copulas

An n-dim. copula is the joint distribution function $C : [0,1]^n \to [0,1]$ of an n-dim. random vector $(U_1, \ldots, U_n)$, whose components U_i are uniformly distributed on $[0,1]$

Vine copulas

Multivariate copulas

An n-dim. copula is the joint distribution function $C : [0,1]^n \to [0,1]$ of an n-dim. random vector $(U_1, \ldots, U_n)$, whose components U_i are uniformly distributed on $[0,1]$

Vine copula

A vine copula is obtained by decomposing a multivariate density into conditional densities and applying Sklar's theorem sequentially to each conditional density

Vine copulas

Multivariate copulas

An n-dim. copula is the joint distribution function $C : [0,1]^n \to [0,1]$ of an n-dim. random vector $(U_1, \ldots, U_n)$, whose components U_i are uniformly distributed on $[0,1]$

Vine copula

A vine copula is obtained by decomposing a multivariate density into conditional densities and applying Sklar's theorem sequentially to each conditional density

Vine copulas as families of trees

We can interpret a vine copula as a family of trees, where

Vine copulas

Multivariate copulas

An n-dim. copula is the joint distribution function $C : [0,1]^n \to [0,1]$ of an n-dim. random vector $(U_1, \dots, U_n)$, whose components U_i are uniformly distributed on $[0,1]$

Vine copula

A vine copula is obtained by decomposing a multivariate density into conditional densities and applying Sklar's theorem sequentially to each conditional density

Vine copulas as families of trees

We can interpret a vine copula as a family of trees, where

- each edge is a conditional bivariate copula

Vine copulas

Multivariate copulas

An n-dim. copula is the joint distribution function $C : [0,1]^n \to [0,1]$ of an n-dim. random vector $(U_1, \ldots, U_n)$, whose components U_i are uniformly distributed on $[0,1]$

Vine copula

A vine copula is obtained by decomposing a multivariate density into conditional densities and applying Sklar's theorem sequentially to each conditional density

Vine copulas as families of trees

We can interpret a vine copula as a family of trees, where

- each edge is a conditional bivariate copula
- each node is a conditional cumulative distribution function

R-vine

A regular vine (short: R-vine) V on n elements is a family of trees $\{\mathcal{T}_1, \ldots, \mathcal{T}_{n-1}\}$ with edges $E(V) = E_1 \cup \ldots \cup E_{n-1}$, such that

1. $\mathcal{T}_1 = (N_1, E_1)$ is a connected tree with nodes $N_1 = \{1, \ldots, n\}$ and edges E_1
2. $\mathcal{T}_k$ is a tree with nods $N_k = E_{k-1}$ for all $k \in \{2, \ldots, n-1\}$
3. $\#(e_1 \Delta e_2) = 2$ for all $\{e_1, e_2\} \in E_k$ with $k \in \{2, \ldots, n-1\}$

R-vine

A regular vine (short: R-vine) V on n elements is a family of trees $\{\mathcal{T}_1, \ldots, \mathcal{T}_{n-1}\}$ with edges $E(V) = E_1 \cup \ldots \cup E_{n-1}$, such that

1. $\mathcal{T}_1 = (N_1, E_1)$ is a connected tree with nodes $N_1 = \{1, \ldots, n\}$ and edges E_1
2. $\mathcal{T}_k$ is a tree with nods $N_k = E_{k-1}$ for all $k \in \{2, \ldots, n-1\}$
3. $\#(e_1 \Delta e_2) = 2$ for all $\{e_1, e_2\} \in E_k$ with $k \in \{2, \ldots, n-1\}$

R-vine decomposition

The decomposition of a n-dim. density $f_{1,\ldots,n}$ corresponding to an R-vine V with edges $E(V)$ is given by

$$f_{1,\ldots,n} = \prod_{e \in E(V)} c_{t_1,t_2|S(e)}(F_{t_1|S(e)}, F_{t_2|S(e)}) \cdot \prod_{j=1}^{n} f_j,$$

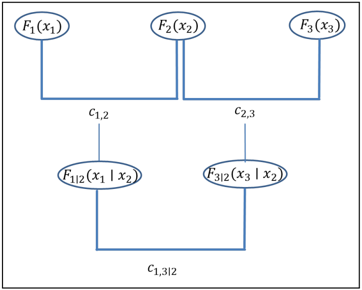
where $S(e)$ is the so-called conditioning set, $T(e) = \{t_1, t_2\}$ is the conditioned set of the edge e and f_j are the one-dim. marginal densities

Outline

D-vine copulas

D-vines

- D-vines are a special type of R-vine

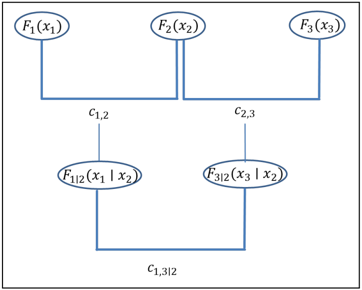


D-vine structure for 3 random variables

D-vine copulas

D-vines

- D-vines are a special type of R-vine
- Each node is connected to not more than 2 edges

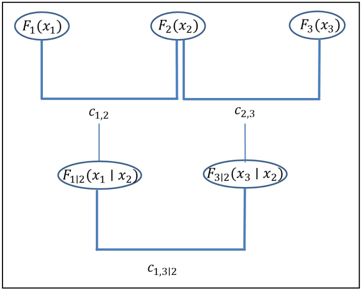


D-vine structure for 3 random variables

D-vine copulas

D-vines

- D-vines are a special type of R-vine
- Each node is connected to not more than 2 edges
- For each edge in $E_k = \{e_1, \ldots, e_{n-k}\}$ the conditioned set is $T(e_i) = \{i, i+k\}$ and the conditioning set is $S(e_i) = \{i+1, \ldots, i+k-1\}$

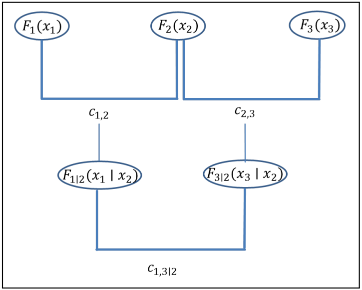


D-vine structure for 3 random variables

D-vine copulas

D-vines

- D-vines are a special type of R-vine
- Each node is connected to not more than 2 edges
- For each edge in $E_k = \{e_1, \ldots, e_{n-k}\}$ the conditioned set is $T(e_i) = \{i, i+k\}$ and the conditioning set is $S(e_i) = \{i+1, \ldots, i+k-1\}$

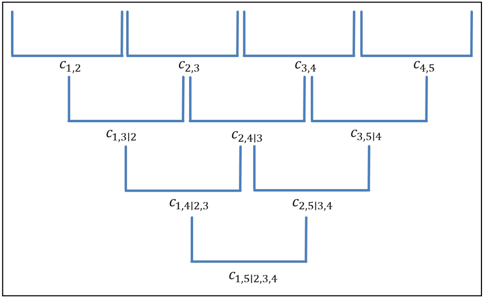


D-vine structure for 3 random variables

For 3-dim. densities a D-vine corresponds to following decomposition:

$$\begin{aligned} f_{1,2,3}(x_1, x_2, x_3) &= f_{3|1,2}(x_3 \mid x_1, x_2) f_{2|1}(x_2 \mid x_1) f_1(x_1) \\ &= c_{1,3|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)) f_{3|2}(x_3 \mid x_2) f_{2|1}(x_2 \mid x_1) f_1(x_1) \\ &= c_{1,3|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)) c_{2,3}(F_2(x_2), F_3(x_3)) f_3(x_3) \\ &\quad c_{1,2}(F_1(x_1), F_2(x_2)) f_2(x_2) f_1(x_1) \end{aligned}$$

D-vine copulas



D-vine structure for 5 random variables

The decomposition of an *n*-dim. density corresponding to D-vines is:

$$f_{1,\dots,n}(x_1,\dots,x_n) = \prod_{k=1}^{n-1}\prod_{i=1}^{n-k} c_{i,i+k|i+1,\dots,i+k-1}\big(F_{i|i+1,\dots,i+k-1}(x_i \mid x_{i+1},\dots,x_{i+k-1}),$$

$$F_{i+k|i+1,\dots,i+k-1}(x_{i+k} \mid x_{i+1},\dots,x_{i+k-1})\big) \cdot \prod_{j=1}^{n} f_j(x_j)$$

Outline

Fitting of vine copulas

Sequential estimation

The following steps are applied, starting with the first row:

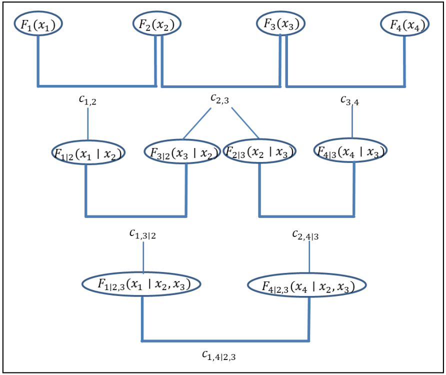
Step 0: Fit the marginal cdfs

Step 1: Transform the data based on the computed cdfs

Step 2: Fit bivariate copulas to the transformed data

Step 3: Compute conditional cdfs using the bivariate copulas

Step 4: Repeat Step 1-3 till the end



Fitting a D-vine with 4 random variables

Fitting of vine copulas

Sequential estimation

The following steps are applied, starting with the first row:

Step 0: Fit the marginal cdfs

Step 1: Transform the data based on the computed cdfs

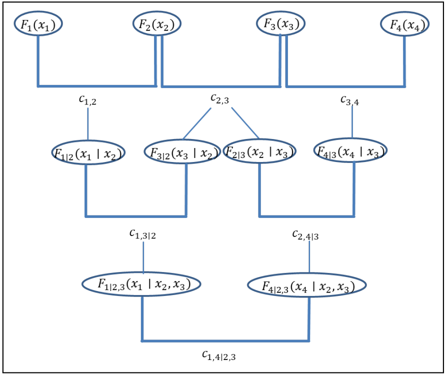
Step 2: Fit bivariate copulas to the transformed data

Step 3: Compute conditional cdfs using the bivariate copulas

Step 4: Repeat Step 1-3 till the end

Fitting of bivariate copulas

Apply ML estimation to fit one-parametric Archimedean copulas



Fitting a D-vine with 4 random variables

Outline

Compute conditional probabilities

Goal

For a predefined threshold v and feed-in points $p_1, \ldots, p_n$ compute the conditional probabilities $P(S_1 + \ldots + S_n \geq v \mid R_1 = r(p_1, t), \ldots, R_n = r(p_n, t))$ given global radiation forecasts $r(p_1, t), \ldots, r(p_n, t)$ and forecast time t

Application of D-vine copulas

Fit an $n+1$-dim. D-vine copula to the random vector $(R_1, \ldots, R_n, S_1 + \ldots + S_n)$

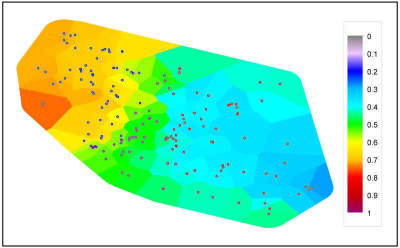
Computation of conditional probabilities

Based on the fitted D-vine copula we compute

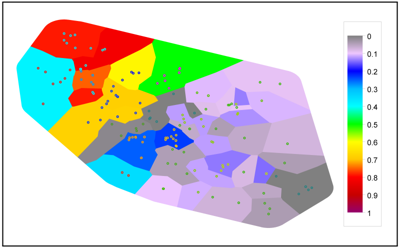
$$P(S_1 + \ldots + S_n \geq v \mid R_1 = r(p_1, t), \ldots, R_n = r(p_n, t)) =$$

$$\int_v^1 c_{1,n+1|2,\ldots,n}(F_{1|2,\ldots,n}(r(p_1,t) \mid r(p_2,t), \ldots, r(p_n,t)), F_{n+1|2,\ldots,n}(s \mid r(p_1,t), \ldots, r(p_{n-1},t)))\mathrm{d}s$$

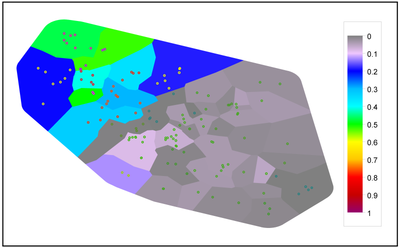
Conditional probabilities calculated by multivariate D-vines



Normalized global radiation forecast

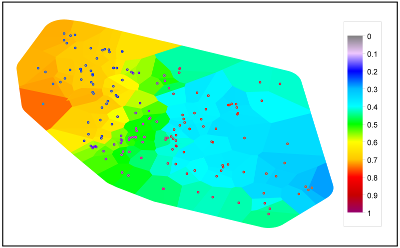


Conditional probabilities for threshold $v = 0.7$

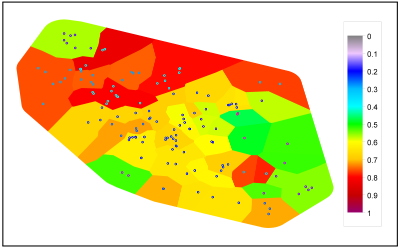


Conditional probabilities for threshold $v = 0.8$

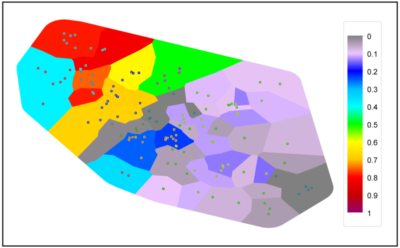
Conditional probabilities calculated by multivariate D-vines



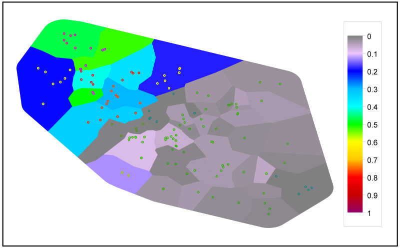
Normalized global radiation forecast



Normalized aggregated solar power supply



Conditional probabilities for threshold $v = 0.7$



Conditional probabilities for threshold $v = 0.8$

Literature

- Karimi, M., Mokhlis, H., Naidu, K., Uddin, S. and Bakar, A.H.A., 2016. Photovoltaic penetration issues and impacts in distribution network - A review. Renewable and Sustainable Energy Reviews, 53, 594-605.
- Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 1-22.
- Nelsen, R.B., 2006. An Introduction to Copulas. Springer.
- Joe, H., 2014. Dependence Modeling with Copulas. Chapman and Hall/CRC.
- von Loeper, F., Schaumann, P., de Langlard, M., Hess, R., Bäsmann, R. and Schmidt, V., 2019. Probabilistic prediction of solar power supply to distribution networks, using forecasts of global radiation. Preprint (submitted)