Probabilistic Forecasting of Electricity Demand using Markov Chain and Statistical Distribution

Eralda Gjika, Aurora Ferrja, Lule Basha,

Department of Applied Mathematics,
Faculty of Natural Science, University of Tirana
Tirana, Albania

E-Mail: eralda.dhamo@fshn.edu.al

Arbesa Kamberi

Albanian Power Corporate
Tirana, Albania

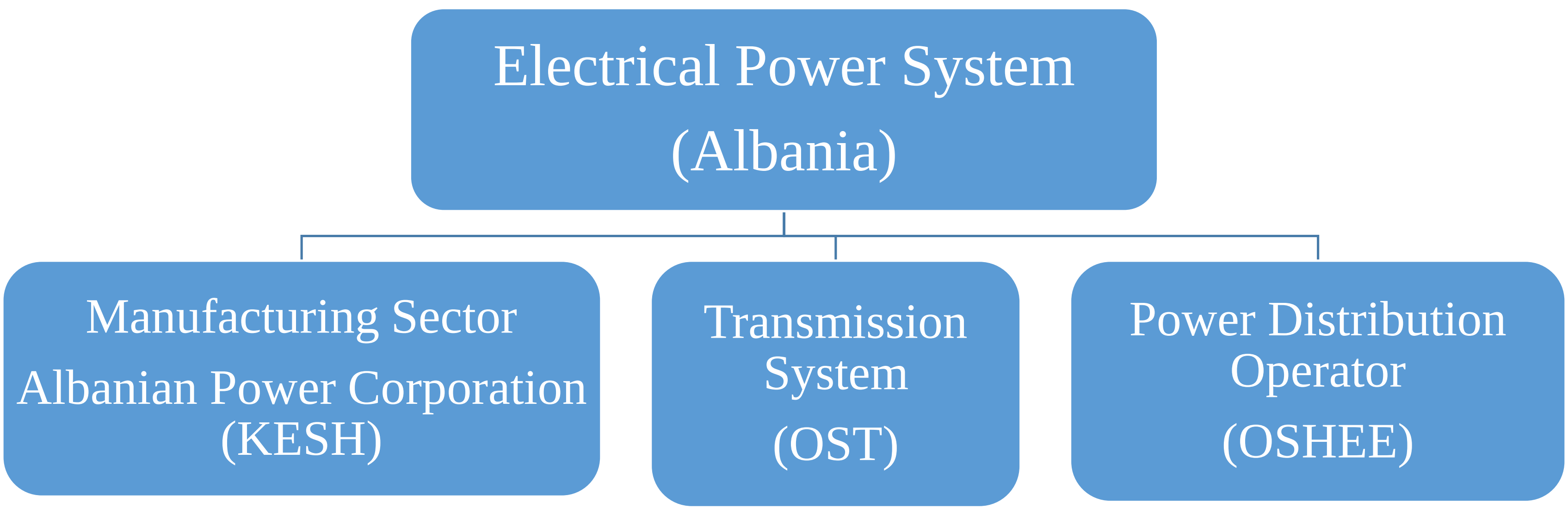
E-Mail: kamberia@kesh.al

Electrical Energy Power system in Albania

The Mediterranean geographical position and climatic conditions of Albania makes the power sector heavily dependent on electrical energy produced mainly by **hydropower plants (HPP).**

The electrical power system is divided into three main sectors:

- Manufacturing sector
- Transmission
- Distribution

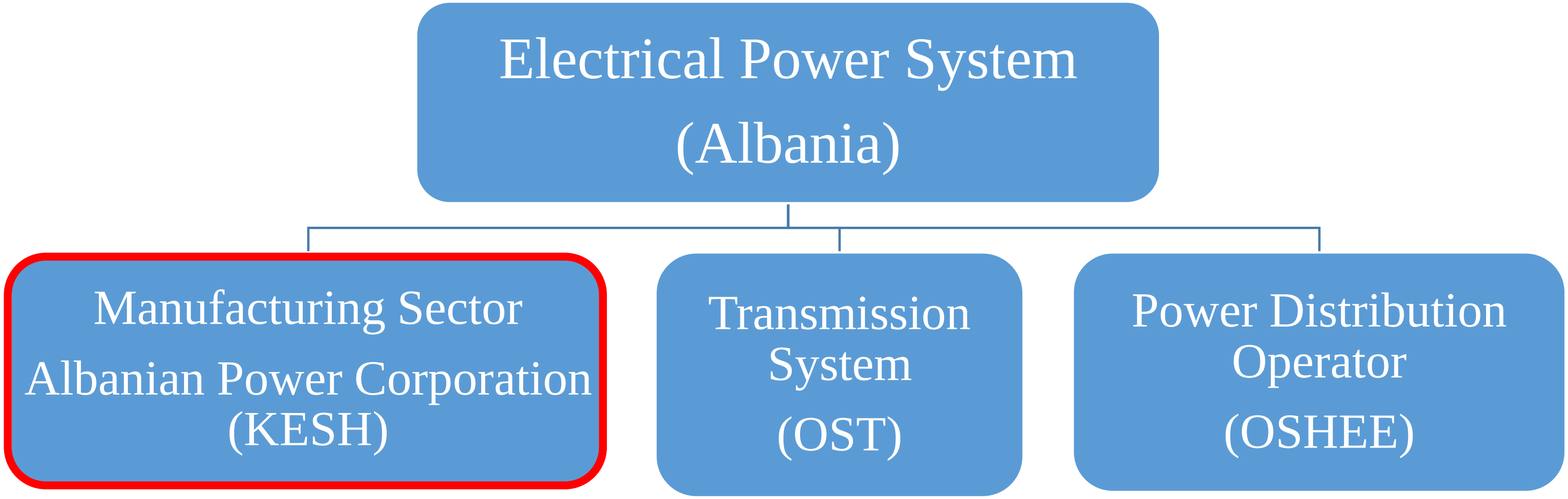


Electrical Energy Power system in Albania

The Mediterranean geographical position and climatic conditions of Albania makes the power sector heavily dependent on electrical energy produced mainly by **hydropower plants (HPP).**

The electrical power system is divided into three main sectors:

- Manufacturing sector
- Transmission
- Distribution



Manufacturing Sector
Albanian Power Corporation (KESH)

KESH is the main public producer of electrical energy in the country. It has into administration the main HPP positioned in Drin Cascade (Fierza HPP, Koman HPP, Vau-Dejes HPP) with an installed capacity of **1,350 MW.**

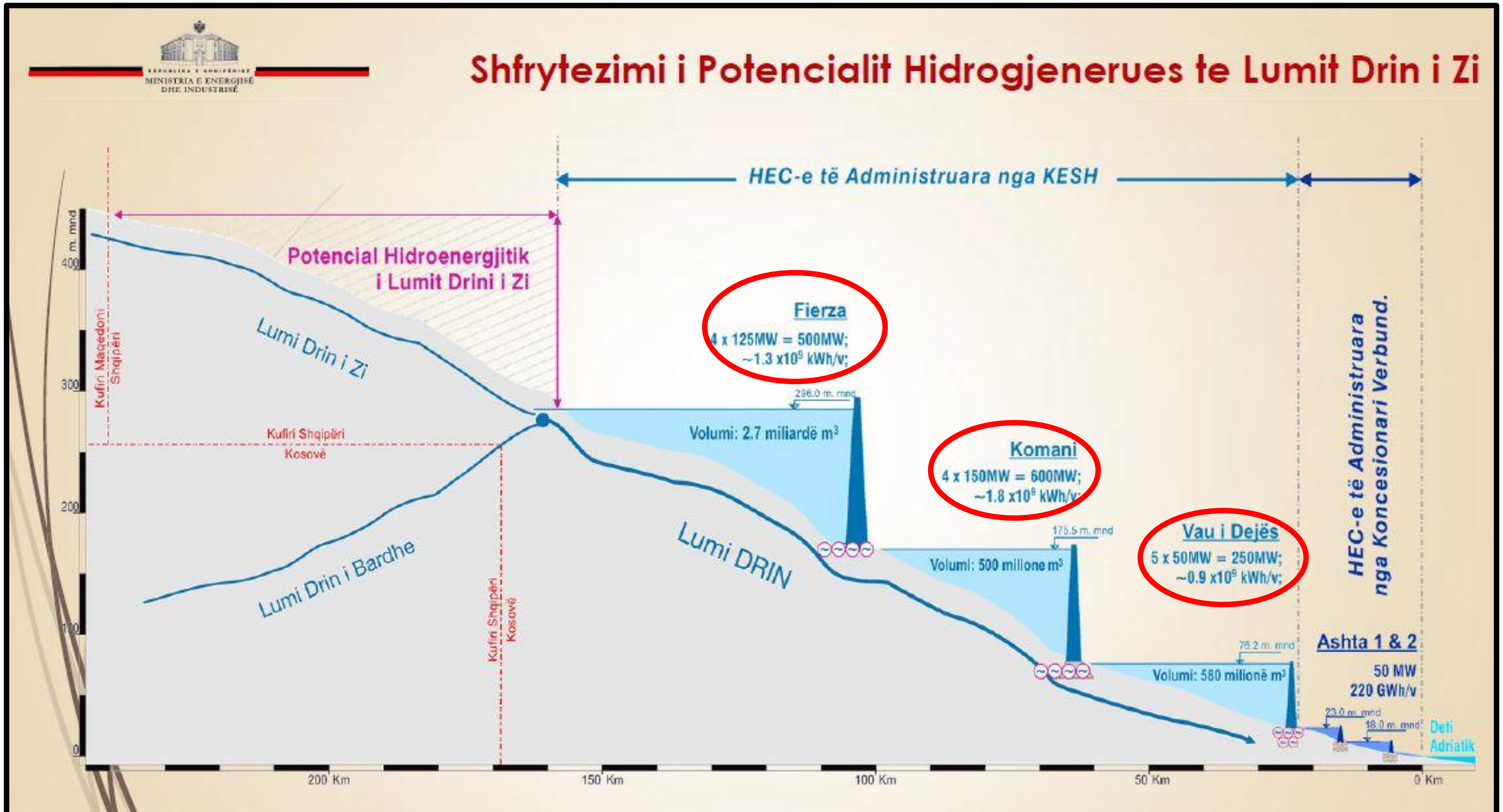
The cascade built on the Drin River Basin is the **largest in the Balkan** both for its installed capacity and the size of hydro-tech works. Having in operation **79% of production** capacity in the country, KESH supplies about **70-75% of the demand for electricity.**

KESH is not only one of the producers of electricity from important hydropower sources in the region, but is also considered a regionally influential factor in the safety of hydro cycles.

http://kesh.al/info.aspx?_NKatID=1211

Position of HPP in Drin Cascade

Fierza is the upper HPP of the Drin river cascade. Then Koman and Vau-Dejes which produce the main amount of energy in Drin Cascade.



Drin Cascade- Fierza HPP

Fierza is the upper HPP of the Drin river cascade. For the installed power, the position and volume of the reservoir, **Fierza plays a key role** in the utilization, regulation and security of the entire cascade.

Why probabilistic forecast?

- In our previous works on energy demand monthly data we have used:

• Classical time series (2015a)
• Particle Swarm Optimization (PSO) models combined with the forecast obtained from time series models and other constraint. (2015)
• Hybrid time series models (SARIMA, ETS, Neural Networks etc.) (2018)

• The most accurate models were

Time series models combined with **PSO**

Why probabilistic forecast?

- Now we have to deal with daily data (from 2011-2018).
- Quantify **uncertainty** in the prediction
- **Optimize** decision making in electricity production (how to spread the production in HPP of the cascade)
- **Maximize** sharpness of predictive distributions subject to calibration

Our goal

- **Find** a predictive probability density function (PDF) which better fits our data
- **Generate** a forecast for electricity energy demand on each HPP

Drin Cascade- FIERZA (1st HPP)

To achieve an accurate forecast of daily electricity energy demand in the country we have worked on the daily observations (from 2011 to 2018) in three main HPP on Drin cascade (**Fierza – Koman -Vau Dejes**).

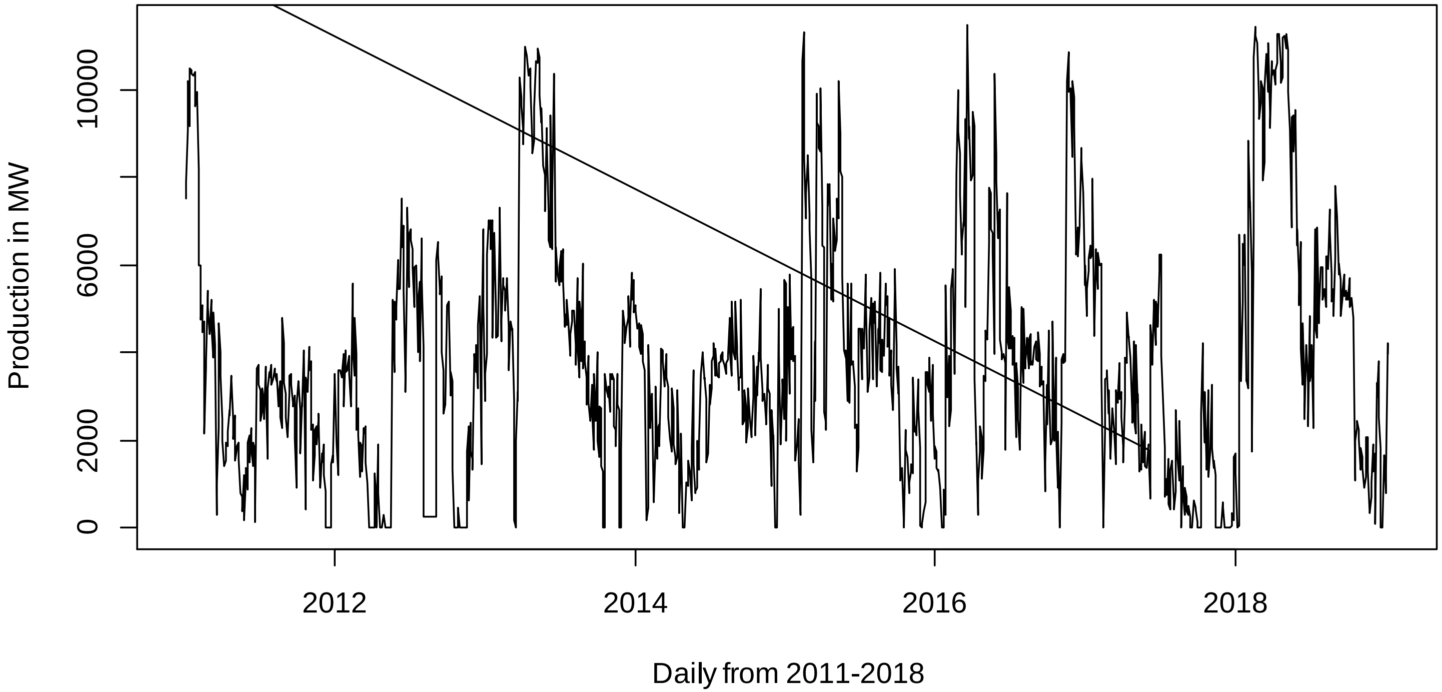


Fig. 1 Daily electricity demand on FIERZA (**1st** HPP)

Drin Cascade- FIERZA (1st HPP)

To achieve an accurate forecast of daily electricity energy demand in the country we have worked on the daily observations (from 2011 to 2018) in three main HPP on Drin cascade (**Fierza – Koman -Vau Dejes**).

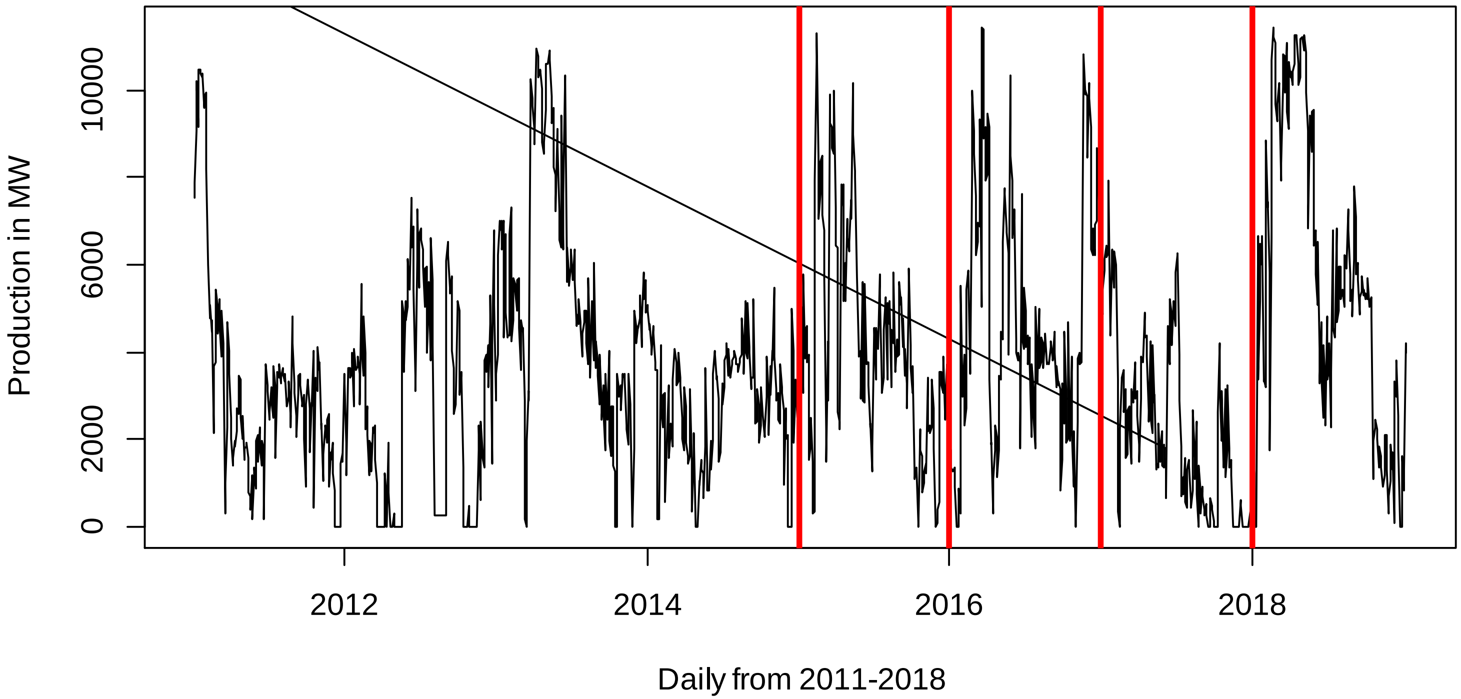


Fig. 1 Daily electricity demand on FIERZA (**1st** HPP)

Drin Cascade- KOMAN (2nd HPP)

To achieve an accurate forecast of daily electricity energy demand in the country we have worked on the daily observations (from 2011 to 2018) in three main HPP on Drin cascade (**Fierza-Koman-Vau Dejes**).

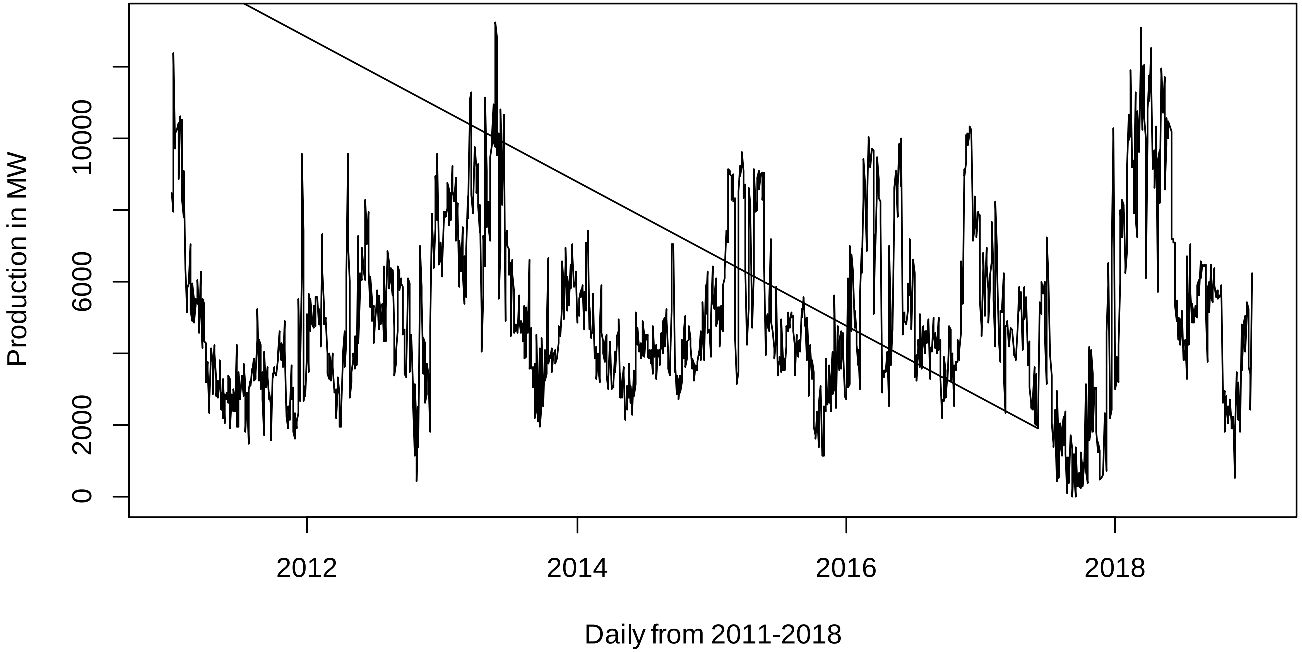


Fig. 2 Daily electricity demand on KOMAN (2nd HPP)

Drin Cascade- KOMAN (2nd HPP)

To achieve an accurate forecast of daily electricity energy demand in the country we have worked on the daily observations (from 2011 to 2018) in three main HPP on Drin cascade (**Fierza-Koman-Vau Dejes**).

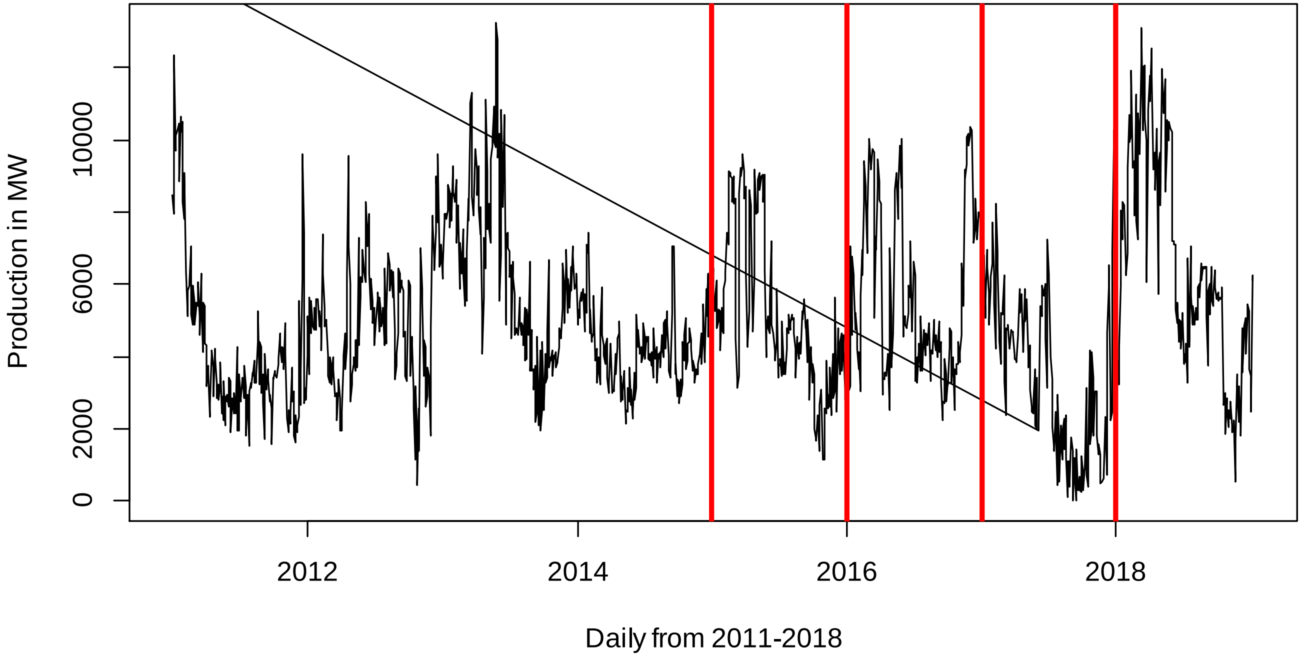


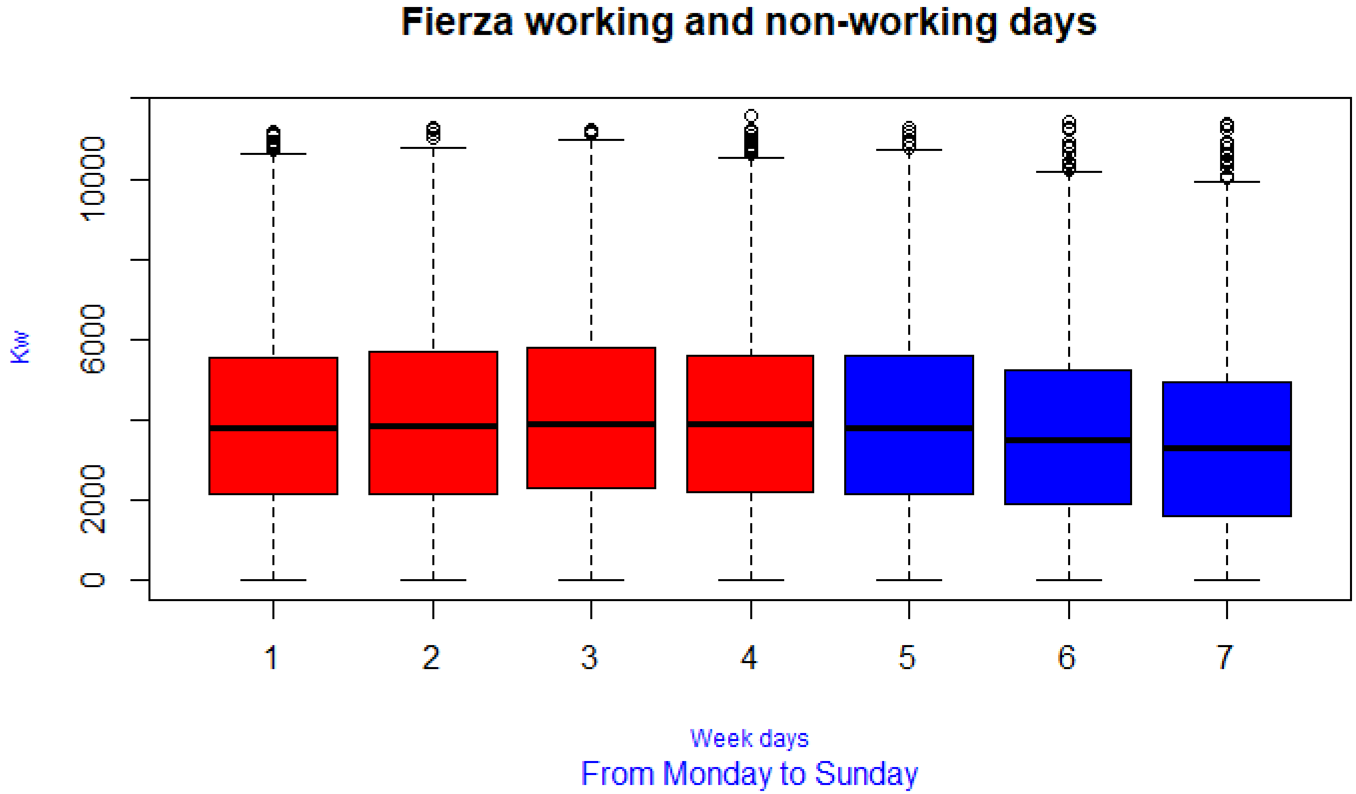
Fig. 2 Daily electricity demand on KOMAN (**2nd** HPP)

Electric Energy Demand differs on working and non-working days?

Working days (official: 8.00-17.00) : Monday to Thursday
Non-working days (official: 8.00-14.00): Friday to Sunday

There is no evidence of a significant difference on demand between working and non-working days for Fierza HPP.

FIERZA

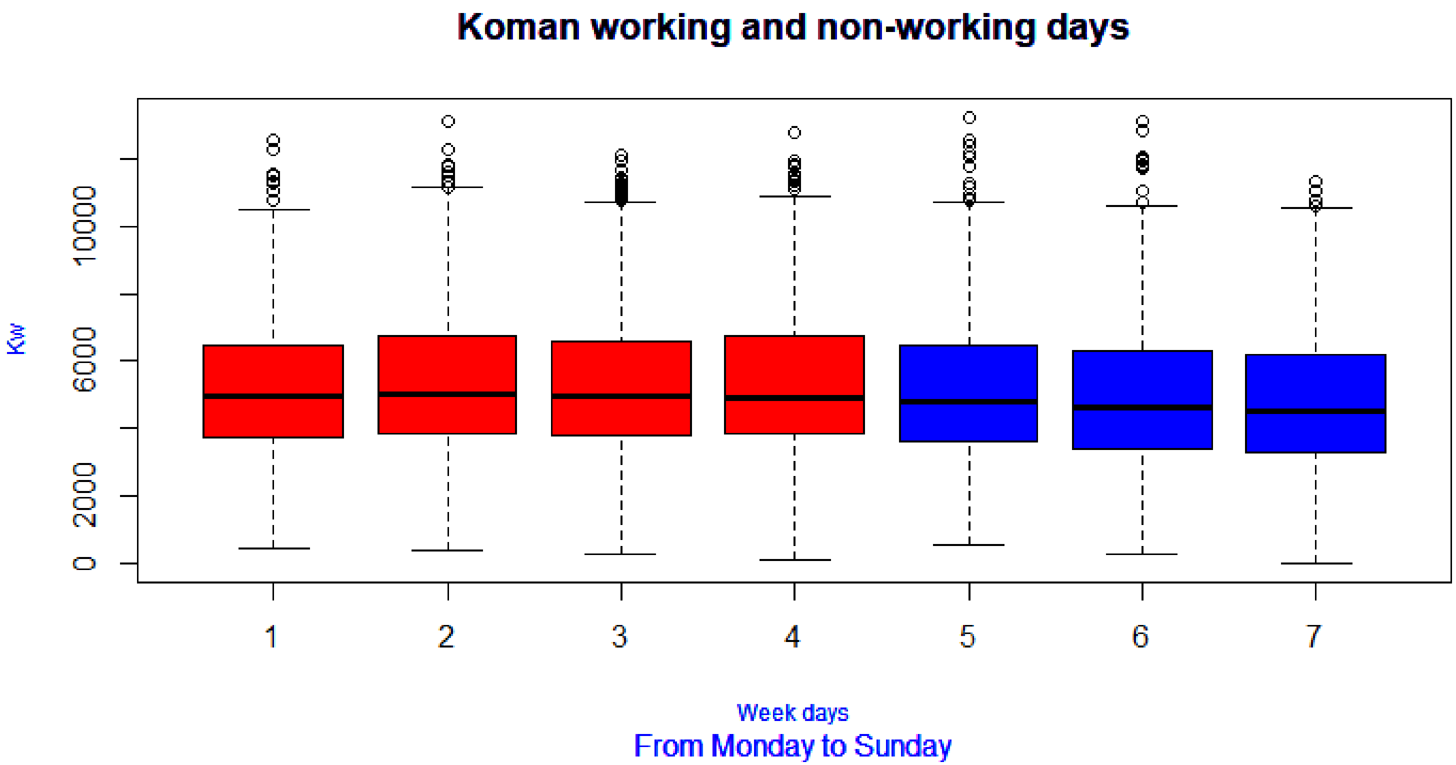


Electric Energy Demand differs on working and non-working days?

Working days (official: 8.00-17.00) : Monday to Thursday
Non-working days (official: 8.00-14.00): Friday to Sunday

There is no evidence of a significant difference on demand between working and non-working days for Koman HPP.

KOMAN



The density histograms of production in Fierza and Koman

Total production | Working days production | Weekend days production

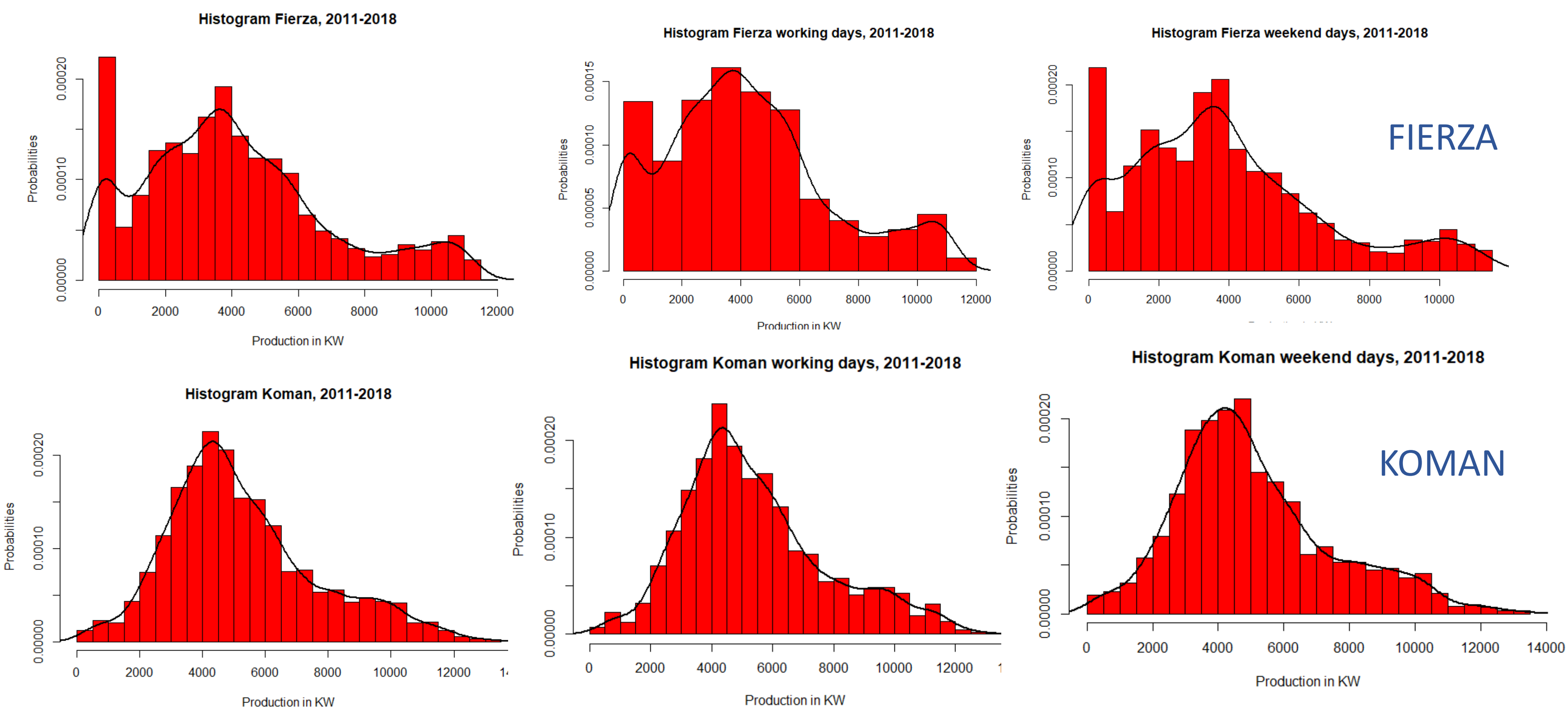


Fig. 3 Density histogram of electricity produced (total-working days-weekend)

Is there a relation among energy produced in the two HPP?

- When the production in Fierza is low then the production in Koman is high. They are used as substitute for each other but, there are other factors affecting the production such as: *precipitations*, *water inflow*, *remount work on HPP* etc.

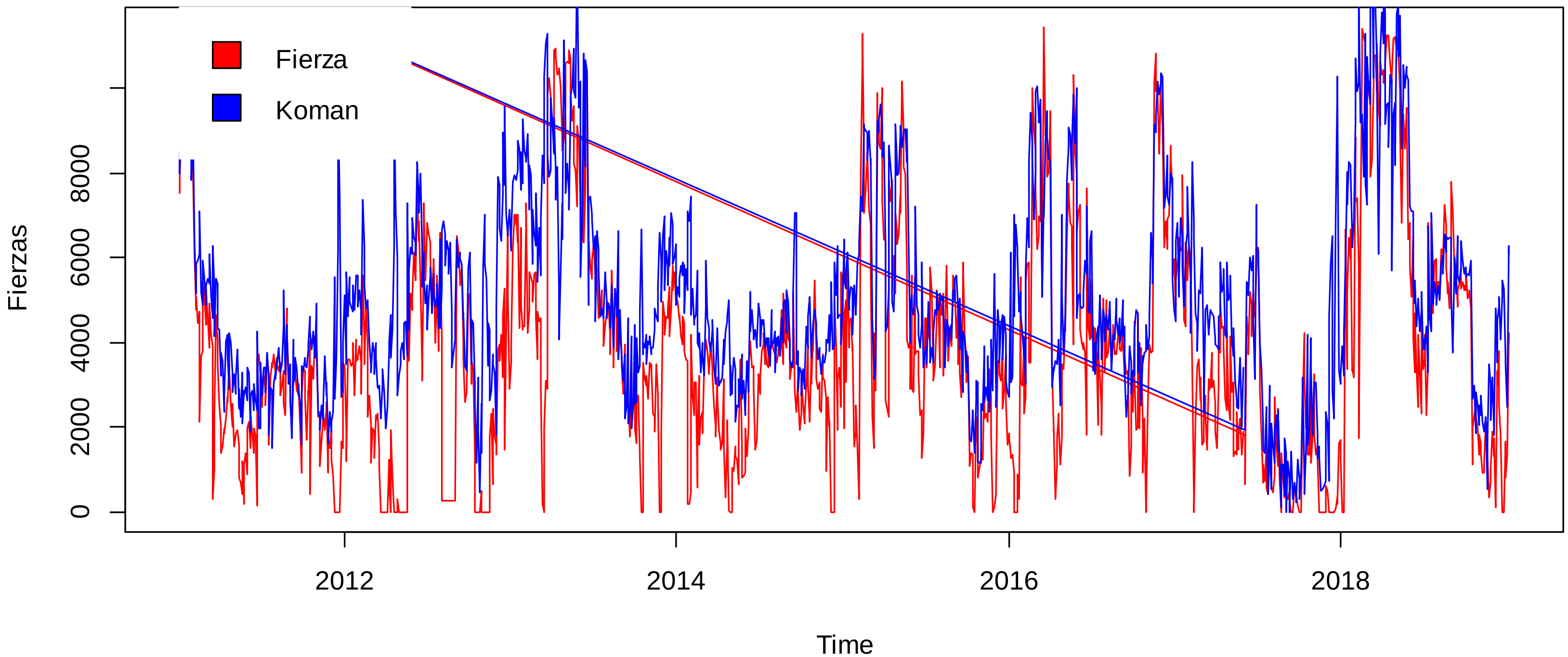


Fig. 4 Relation among production in Fierza and Koman

Is there a relation among the two HPP?

- When the production in Fierza is low then the production in Koman is high. They are used as substitute for each other but, there are other factors affecting the production such as: *precipitations*, *water inflow*, *remount work on HPP* etc.

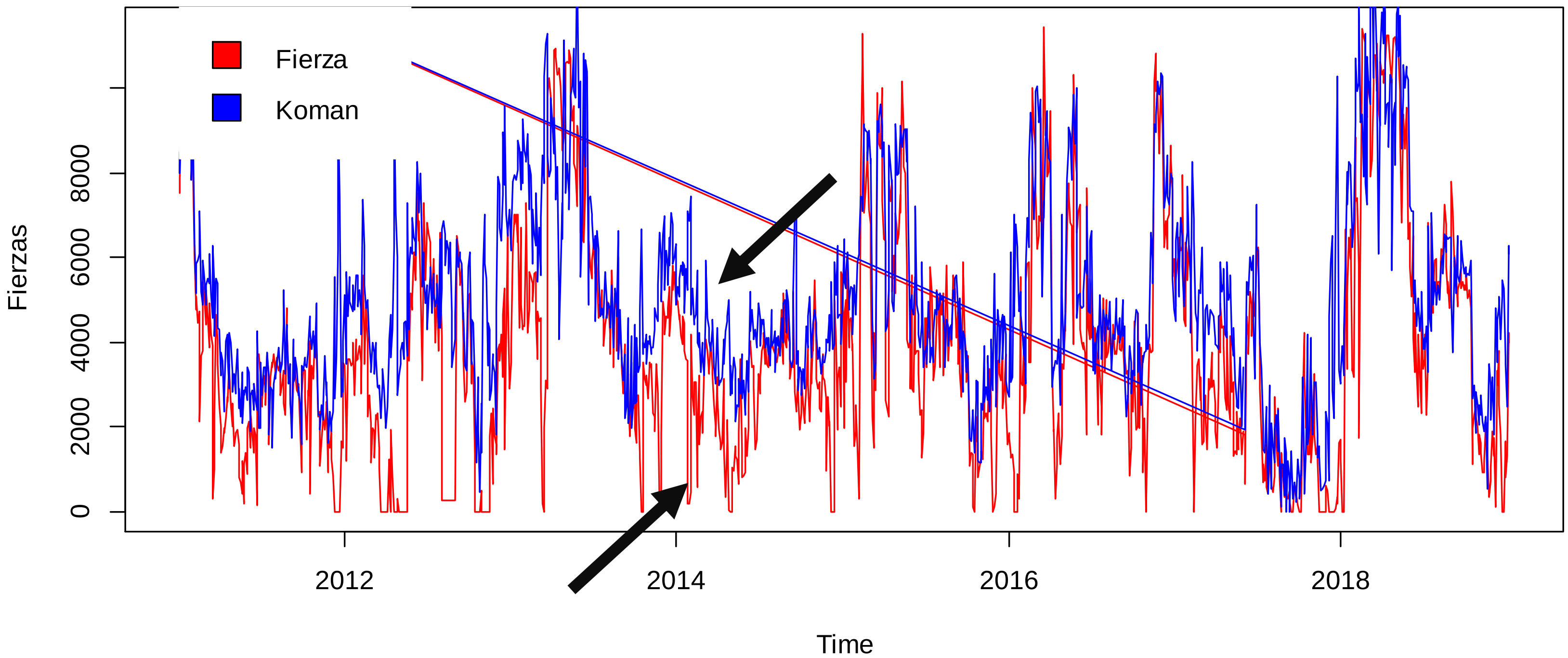


Fig. 4 Relation among production in Fierza and Koman

Is there a correlation among the two HPP?

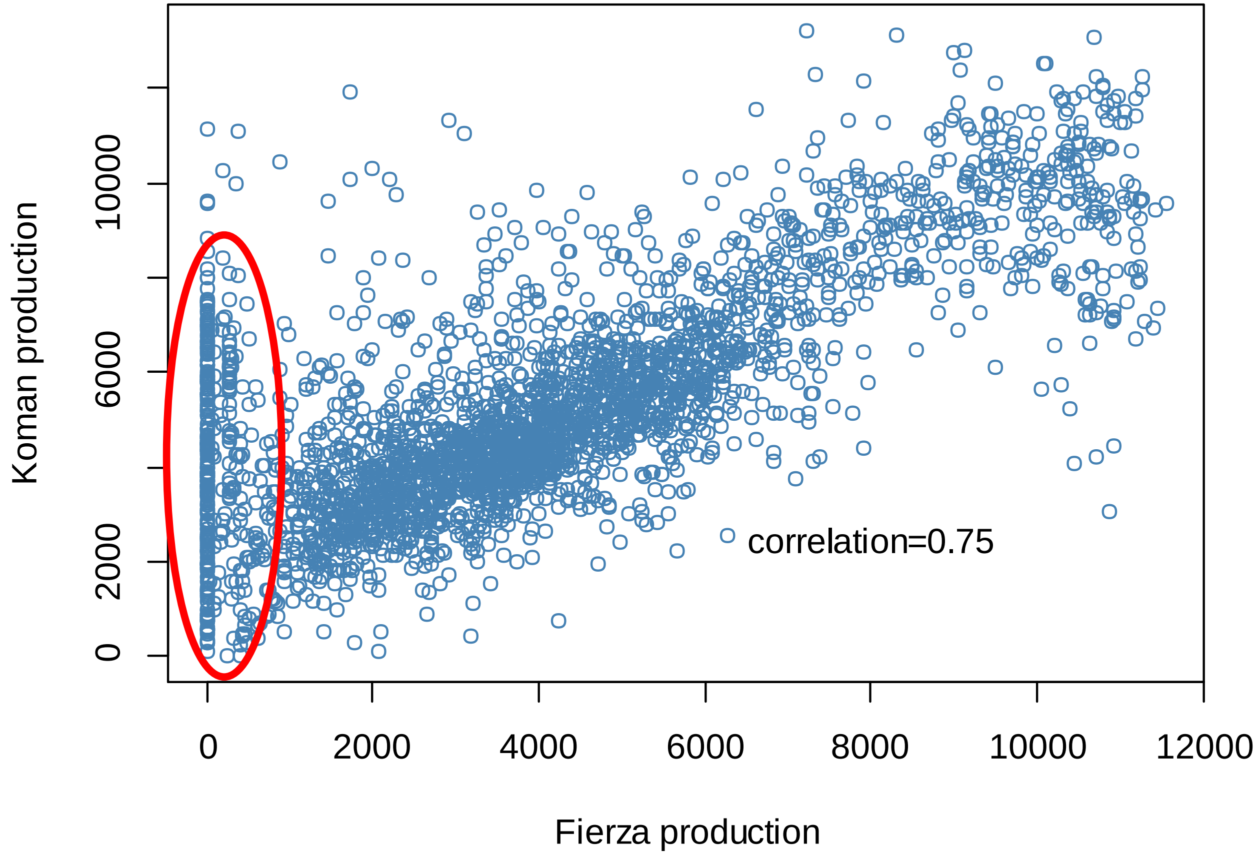


Fig. 5 Correlation among daily energy produced in Fierza and Koman

Why MCMC?

- Monte Carlo Markov Chain (MCMC) is designed to construct an ergodic Markov chain with a distribution *f* as its stationary distribution.
- Asymptotically the chain will resemble samples from *f*.
- A very powerful property of MCMC is that it is possible to combine several samplers into mixtures and cycles of the individual samplers (Tierney, 1994)
- Given that our data for Fierza show a bimodal normal distribution we have used MCMC simulation to fit the parameters of the distributions.
- And, then use it as a probability distribution function to predict the probability of demand being in one of the intervals.

MCMC Simulations

Metropolis-Hasting Algorithm

- Metropolis and Ulam (1949), Metropolis et al. (1953) and Hastings (1970) where the first who propose the MCMC procedure.
- All other MCMC models are modification of the base model proposed by Metropolis-Hastings.

- The goal of M-H algorithm is to draw samples from some distribution $p(\theta)$ where $p(\theta) = f(\theta)/K$, where the normalizing constant K may not be known, and very difficult to compute.

- R-Packages : mixtools, mixdist,

Ergodic MC provides an effective algorithm for sampling from π

- Chain is *irreductible* if:

$$(1.1) \qquad \forall x, y \in \Omega \;\; \exists t > 0 \;\; \text{for which} \;\; P_x^t(y) > 0$$

- *P is aperiodic if :*

$$(1.2) \qquad \forall x, y \in \Omega \;\; \text{we have} \;\; \gcd\{t : P_x^t(y) > 0\} = 1$$

Fundamental Theorem: If P is irreducible and aperiodic, then it is ergodic, i.e

$$(1.3) \; \forall x, y \in \Omega \quad \text{we have} \quad P_x^t(y) \xrightarrow[t \to \infty]{} \pi(y)$$

where π is the (unique) stationary distribution of P – i.e $\pi P = \pi$.

Markov Chain procedure on Fierza

- Discretization

We have used the arithmetic mean as a divider among the states {1,2,3,4} which are the levels of the demand respectively:

Low, Lower-Medium, Upper-Medium, High

$$states = \begin{cases} 1 \quad \text{if} \quad \min(X) \le x_i < m_1 \\ 2 \quad \text{if} \quad m_1 \le x_i < m_2 \\ 3 \quad \text{if} \quad m_2 \le x_i < m_3 \\ 4 \quad \text{if} \quad m_3 \le x_i < \max(X) \end{cases}$$

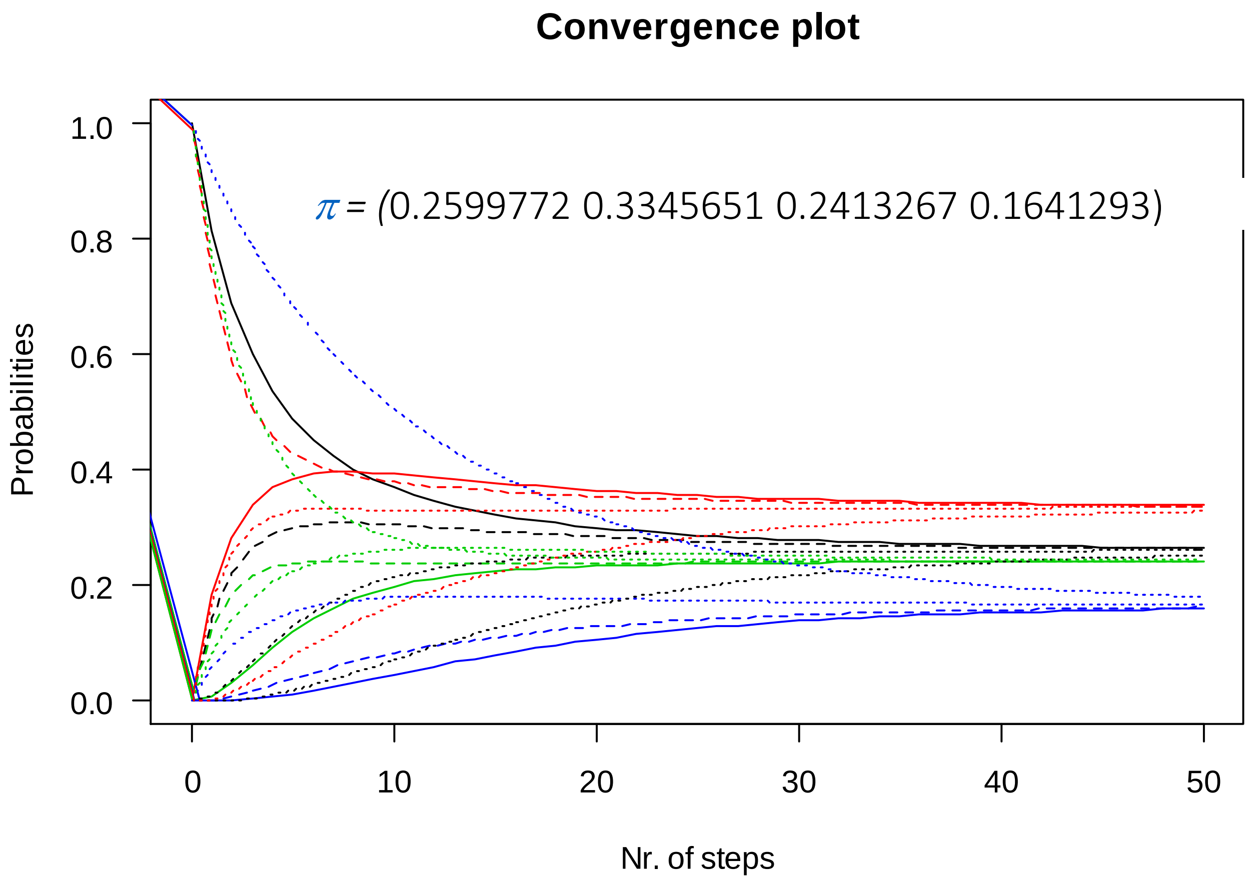
$$m_1 = mean(X[\min, m_2]; m_2 = mean(X); m_3 = mean(X[m_2, \max(X)];$$

- One step ahead transition matrix for Fierza Markov Chain:

$$P_{Fierza} = \begin{pmatrix} 0.813492063 & 0.1812169 & 0.005291005 & 0.00000000 \\ 0.140801644 & 0.7379239 & 0.121274409 & 0.00000000 \\ 0.005698006 & 0.1680912 & 0.770655271 & 0.05555556 \\ 0.000000000 & 0.0000000 & 0.081632653 & 0.91836735 \end{pmatrix}$$

Stationary distribution for Fierza

Convergence probability Plot for Fierza Markov Chain:



MCMC model for Fierza

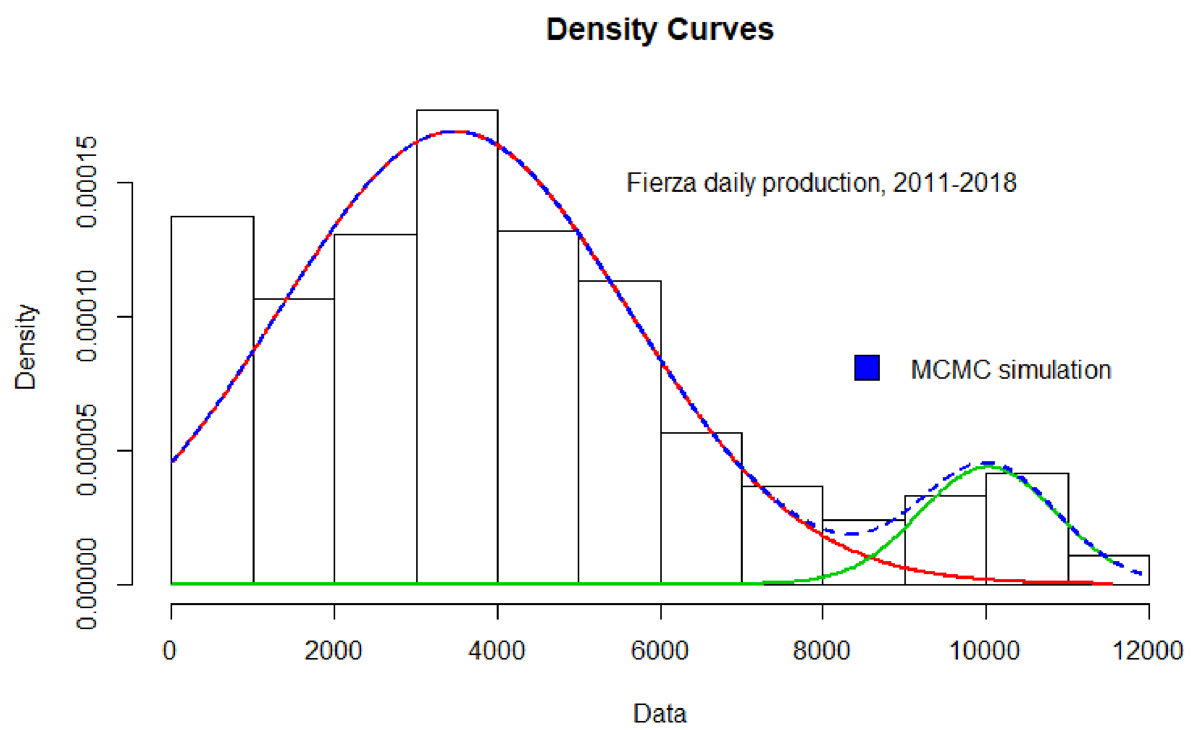
Fierza production can be modeled as a bimodal normal distribution with parameters

$$\mu_1 = 3469.747,\ \sigma_1 = 2141.3638, \quad \text{and} \quad \mu_2 = 10022.759, \sigma_2 = 843.4492$$

$$\lambda = (0.90750601,\ 0.09249399)$$

- With the probability density function (pdf) simulated by MCMC:

(1) $$f_F(x) = \lambda[1] \cdot N(\mu_1, \sigma_1^2) + \lambda[2] \cdot N(\mu_2, \sigma_2^2)$$



MCMC model for Koman

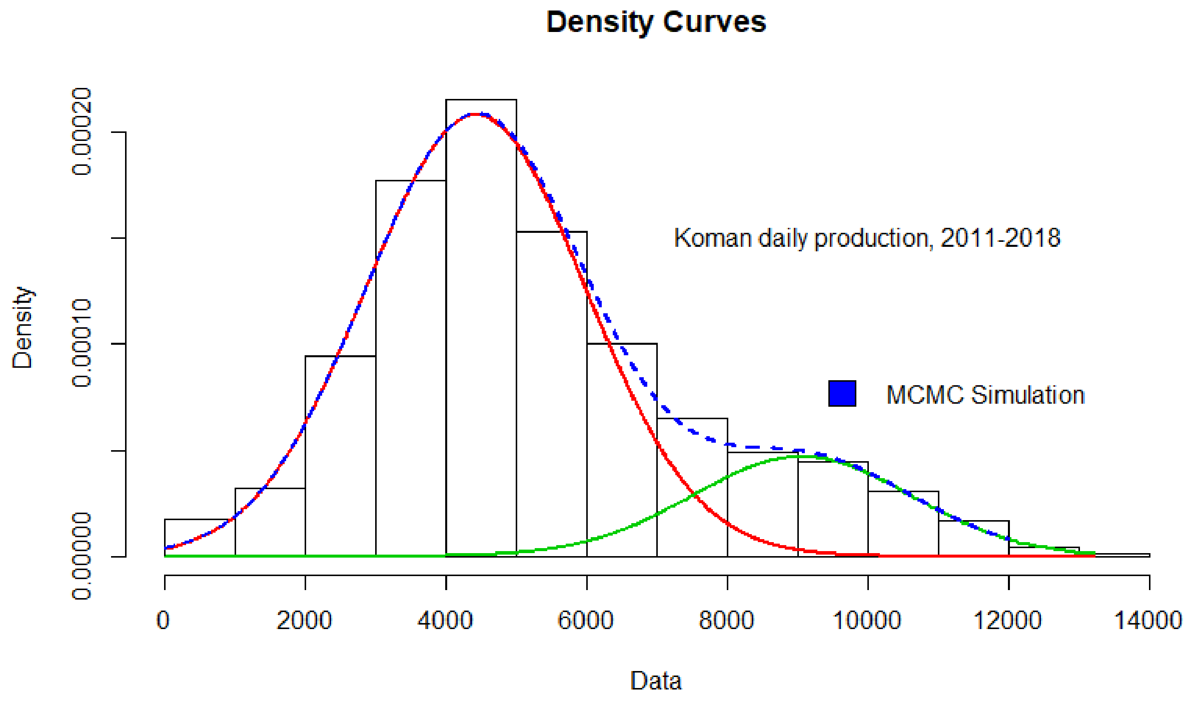
- Koman production can be modeled as a bimodal normal distribution with parameters

$$\mu_1 = 4423.172,\ \sigma_1 = 1565.755 \quad \text{and} \quad \mu_2 = 9053.173, \sigma_2 = 1566.237$$

$$\lambda = (0.8158997,\ 0.1841003)$$

With the probability density function (pdf) simulated by MCMC:

$$f_K(x) = \lambda[1] \cdot N(\mu_1, \sigma_1^2) + \lambda[2] \cdot N(\mu_2, \sigma_2^2)$$



MCMC model for Vau-dejes

- Vau-Dejes production can be modeled as a bimodal normal distri

bution with parameters

$$\mu_1 = 2223.964,\ \sigma_1 = 877.2833 \quad \text{and} \quad \mu_2 = 4903.466, \sigma_2 = 620.3383$$

$$\lambda = (0.822303,\ 0.177617)$$

With the probability density function (pdf) simulated by MCMC:

$$f_V(x) = \lambda[1] \cdot N(\mu_1, \sigma_1^2) + \lambda[2] \cdot N(\mu_2, \sigma_2^2)$$

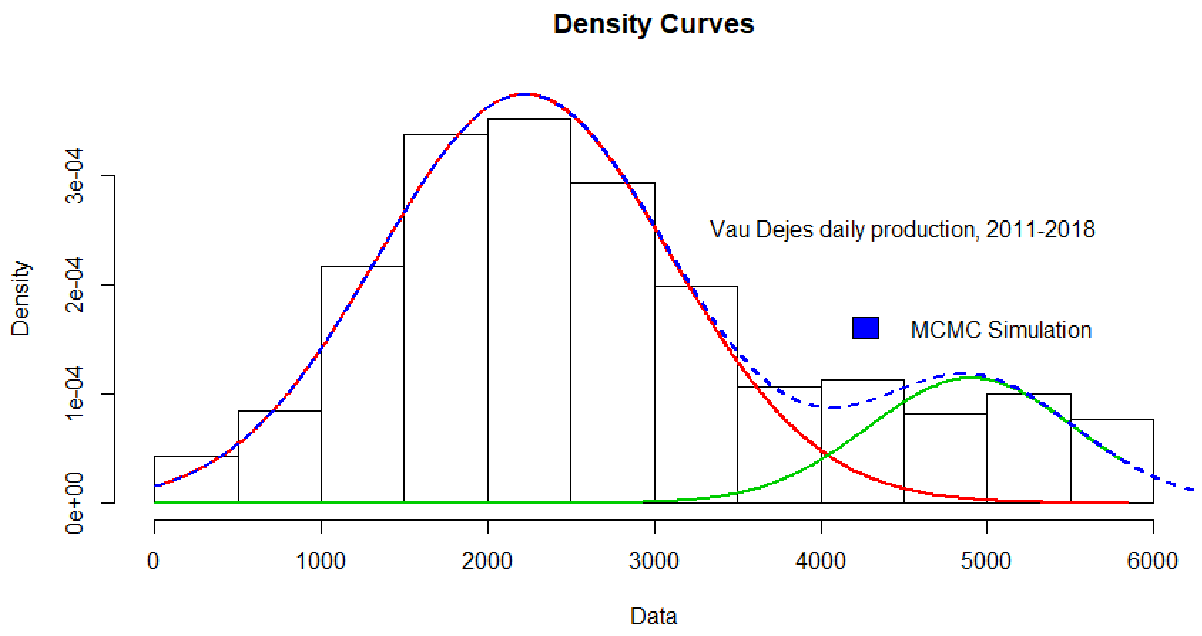


Table 1 Probability estimation of energy produced by each HPP-using MCMC

	Production probability			
Demand (MW)	Fierza	Koman	Vau Dejes	Max probability
0-500	0.02737	0.00306	0.01569	0.02737478
500-1000	0.03779	0.00676	0.04669	0.04669129
1000-1500	0.04941	0.01351	0.10126	0.1012567
1500-2000	0.06118	0.0244	0.16004	0.1600444
2000-2500	0.07177	0.03983	0.1844	0.1844002
2500-3000	0.07973	0.05877	0.15504	0.1550422
3000-3500	0.0839	0.07838	0.09671	0.09671478
3500-4000	0.08363	0.09452	0.0531	0.0945186
4000-4500	0.07895	0.1031	0.04664	0.1030971
4500-5000	0.0706	0.10184	0.0573	0.101844
5000-5500	0.05979	0.09139	0.04855	0.09139356
5500-6000	0.04797	0.07507	0.02309	0.07507371
6000-6500	0.03645	0.05748	0.00596	0.05747539
6500-7000	0.02625	0.04262	0.00083	0.0426218
7000-7500	0.01799	0.03267	0	0.0326673
7500-8000	0.01218	0.0276	0	0.02760282
8000-8500	0.00958	0.02585	0	0.02584838
8500-9000	0.01122	0.02527	0	0.02526543
9000-9500	0.01659	0.0241	0	0.02409558
9500-10000	0.02166	0.0215	0	0.02165692
10000-10500	0.02139	0.01759	0	0.02138559
10500-11000	0.01529	0.01308	0	0.01529351
11000-11500	0.00783	0.00881	0	0.008811653
11500-12000	0.00286	0.00537	0	0.005370822
12000-12500	0.00075	0.00443	0	0.004434634
12500-13000	0.00014		0	0.000141361

Conclusions

- Compared to other statistical distributions (*Weibull*, *gamma*, *normal*, *log-normal*, *Exponentiated G distributions* etc.) we have tried to fit to our data it seems that:

MCMC models were effective on modeling the probability distribution of production in each HPP on Drin cascade.

$$X_t^F \sim N(\mu_1^F, \sigma_1^F) + N(\mu_2^F, \sigma_2^F) \quad \text{with} \quad \lambda_F = (\lambda_1^F, 1-\lambda_1^F)$$

$$X_t^K \sim N(\mu_1^K, \sigma_1^K) + N(\mu_2^K, \sigma_2^K) \quad \text{with} \quad \lambda_K = (\lambda_1^K, 1-\lambda_1^K)$$

$$X_t^V \sim N(\mu_1^V, \sigma_1^V) + N(\mu_2^V, \sigma_2^V) \quad \text{with} \quad \lambda_V = (\lambda_1^V, 1-\lambda_1^V)$$

Further improvements …

- The next goal would be to optimize daily production in Drin cascade by optimizing the various parameters controlling the production process.

- The optimization process is a highly complex task and must be adjusted to find the best combination of all the variables.

- So, we are working on having a machine learning algorithm capable of predicting the production based on multi-dimensional optimization algorithm that will explore which control variables to adjust in order to maximize production.

(precipitations, water inflow, reservoir capacity, probabilities of production in each HPP etc.)

References of previous work

E. Gjika, A. Ferrja, A. Kamberi "A Study on the Efficiency of Hybrid Models in Forecasting Precipitations and Water Inflow Albania Case Study", Advances in Science, Technology and Engineering Systems Journal, vol. 4, no. 1, pp. 302-310 (2019). DOI: 10.25046/aj040129

Simoni A., **Dhamo (Gjika) E**., (2015) Forecasting the maximum power in hydropower plant using PSO, 6th INTERNATIONAL CONFERENCE Information Systems and Technology Innovations: inducting modern business solutions, Tirana, June 5-6, 2015. ISBN: 978-9928-05-199-8 . http://conference.ijsint.org/sites/default/files/Agenda_ISTI2015.pdf

Simoni A, **Dhamo (Gjika) E.,** (2015) Evolutionary Algorithm PSO and Holt-Winters method applied in hydro Power Plants Optimization. STATISTICS, PROBABILITY & NUMERICAL ANALYSIS 2015 METHODS AND APPLICATIONS Conference, Tirana, 5-6 December 2015. **Proceedings** Buletini i Shkencave te Natyres, Botim Special, ISSN 2305-882X https://sites.google.com/a/fshn.edu.al/fshn/home/botim-special

Dhamo (Gjika) E., Simoni A, (2015) Efficiency of Time Series Models on Predicting Water Inflow, Conference: QUAESTI 2015, Slovakia, Volume: 3, Issue 1. (**Conference Proceedings**) **ISSN:** *2453-7144*, **CD ROM ISSN:** *1339-5572*. **ISBN:** *978-80-554-1170-5.*
DOI: 10.18638/quaesti.2015.3.1.225

- Thank you for your attention!

 #SWEEThessaloniki