Statistical and Probabilistic Parameters Used as Reference for Hydrological-Energy Studies

EUR ING Alessandro Calvi[1]

[1] alessandro.calvi84@gmail.com , ENEL Green Power S.p.A, Hydroelectric Design, Hydroelectric Center of Excellence, Torino, Italy

Abstract

Hydropower plant future production estimation is based on hydrological-hydraulic data. The present metholodogy validates the median as the reference parameter to be used for a better interpretation of statistical series, because it is a centered value where duration curves and their complementary curves intersect. Then, it is not affected by extreme events, providing a well representation of the whole dataset.

Keywords: Hydrology, Hydropower plants, Repowering, Energy, Statistics

Statistical and Probabilistic Parameters Used as Reference for Hydrological-Energy Studies

EUR ING Alessandro Calvi

Abstract

Hydropower plant future production estimation is based on hydrological-hydraulic data. The present metholodogy validates the median as the reference parameter to be used for a better interpretation of statistical series, because it is a centered value where duration curves and their complementary curves intersect. Then, it is not affected by extreme events, providing a well representation of the whole dataset.

Keywords: Hydrology, Hydropower plants, Repowering, Energy, Statistics

1. Introduction

Hydrological phenomena such as precipitation, floods, and droughts are inherently random and unpredictable, posing significant challenges in their analysis and modeling. Due to the complex nature of the hydrological system, a complete understanding of these physical processes is still out of reach, and reliable deterministic models remain under development. As a result, statistical methods have become an essential tool for analyzing hydrological data and informing the design of hydraulic and hydroelectric infrastructures, including flood management systems, drainage works, and water resource management strategies. These methods are particularly valuable in translating complex data into actionable insights for engineers, urban planners, and policymakers. A wide range of statistical techniques are employed in hydrology, each serving a different purpose depending on the specific objectives of the analysis. Common applications of statistical methods in hydrology include simulation, forecasting, uncertainty analysis, spatial interpolation, and risk assessment. The effectiveness of these techniques is highly dependent on the availability and quality of data, particularly in regions with sparse observations. This is especially evident in ungauged areas, where hydrological analysis must rely on indirect data or regional models. Statistical approaches in these regions have proven indispensable, providing reliable estimates in the absence of direct measurements.

Many statistical methods used in hydrology rely on the assumption of stationarity, which posits that statistical properties of hydrological variables do not change over time. While this assumption is simple to define, it is often difficult to verify in practice. Nevertheless, stationarity remains a key consideration in many statistical procedures, as violations of this assumption can lead to significant errors in the results. One of the more promising techniques to address this challenge is trend detection in time series data [1], [2], [3], [4], [5], [6], [7], where methods like segmentation are employed to identify shifts or changes in the behavior of hydrological variables over time.

Another crucial aspect of hydrological analysis involves extreme value (EV) theory [8], [9], [10], which is widely used to estimate the likelihood of extreme events such as heavy rainfall or flooding. The correct application of extreme value analysis is essential to avoid significant under- or overestimation of key hydrological parameters, such as rainfall intensity or runoff volumes. Various probability distributions and estimation techniques are used to model the tail of the distribution of hydrological events, with a particular focus on return periods for extreme events. The accurate estimation of extreme events is vital for designing infrastructure that can withstand such phenomena. Precipitation is the most widely monitored hydrological phenomenon, and as such, methods for analyzing rainfall patterns, particularly the intensity-duration-frequency (IDF) approach, have been extensively studied. IDF curves are crucial tools for the design of hydraulic systems, providing estimates of rainfall intensity over specific durations and return periods. Although IDF curves have been in use for nearly a century, they have evolved from simple empirical models to more sophisticated and theoretically sound formulations. Recent research has revealed that traditional IDF methods may underestimate rainfall intensity, especially for large return periods, highlighting the need for updated and advanced estimation techniques to ensure the resilience of infrastructure.

In recent years, the introduction of copula functions [11] has further advanced hydrological analysis by allowing for the joint modeling of multiple hydrological variables. Traditionally, hydrologists have focused on individual parameters, such as flood peaks or rainfall intensity, but the copula approach enables a more comprehensive understanding of the interdependencies between different variables, such as rainfall volume, flood duration, and runoff volume. This approach has gained significant traction in hydrological research, offering a more flexible and accurate way to model complex hydrological relationships and improve the prediction of extreme events.

Finally, the challenge of analyzing hydrological data in ungauged basins, particularly small basins with limited data, remains a significant issue. These regions often lack sufficient monitoring stations, making it difficult to apply traditional hydrological models. In such cases, regional frequency analysis is a valuable statistical method that allows hydrologists to estimate hydrological parameters based on data from similar, gauged areas. This approach has led to the development of statistical techniques tailored to handle the scarcity of data in ungauged regions, offering valuable tools for hydrological analysis in these challenging environments. Statistical methods in hydrology continue to evolve, providing researchers and practitioners with increasingly sophisticated tools for understanding and predicting hydrological phenomena. By improving the accuracy and reliability of these methods, particularly in ungauged or data-scarce regions, these techniques play a crucial role in designing resilient infrastructure, managing water resources, and mitigating the impacts of extreme weather events.

2. Open challenges

Despite over five decades of research in Regional Frequency Flood Analysis (RFFA), significant challenges remain in selecting the most suitable statistical methods for hydrological studies, particularly when addressing issues like prediction in ungauged basins. While some areas, such as the choice and estimation of regional parent distributions, have been well-explored, others remain open and in need of further investigation to enhance the accuracy of regional predictions. One of the key areas of concern is the estimation of the index flood in ungauged basins, a critical step in the regionalization procedure that continues to introduce substantial uncertainty. The scientific community is actively working to refine existing methodologies and to develop more reliable guidelines for selecting the most appropriate methods depending on the specific circumstances.

An important issue that remains is the need for more straightforward, universally applicable parameters that could facilitate easier integration into engineering practice, especially for applications in hydroelectric systems. The lack of such parameters currently complicates the process of applying statistical methods to the analysis of hydrological and energy-related systems, where accurate and accessible data interpretation is essential. Hydropower infrastructure, in particular, requires clear and consistent statistical tools for assessing water availability, flow characteristics, and flood risks to ensure the safe and efficient design of facilities. In this context, it is crucial to define more usable, less complex metrics that can be readily applied by engineers in the field.

In addition, the process of identifying homogeneous regions or pooling groups for flood analysis has long been debated, and while significant progress has been made, it remains a topic of ongoing research. The development of methods that reduce the subjectivity in selecting these groups—such as the application of cluster analysis or unsupervised artificial neural networks—has shown promise, but there is still room for improvement. Furthermore, the potential for geostatistical techniques to aid in statistical regionalization is growing. Techniques like physiographic space-based interpolation (PSBI) and Topological Kriging (Topkriging) offer new possibilities for regionalizing hydrological variables, bypassing the need to define distinct homogeneous regions. These approaches provide continuous representations of variables like T-year floods along stream networks or in the physiographic space, making them particularly useful for ungauged basins. However, even these advanced methods introduce some uncertainty, which still requires careful consideration when applied in practice.

Ultimately, for statistical methods to become more effective and accessible in hydrology, especially in energy-related applications such as hydropower, it is essential to develop simpler, more intuitive parameters that can be easily integrated into professional engineering workflows. Streamlining the selection of statistical methods and making them more universally applicable could significantly improve the reliability of hydrological analyses and facilitate the design of safer, more efficient energy infrastructure.

3. Use of the Median as a Key Parameter in Hydrological and Energy Analyses

The approach used to determine the most suitable parameter for hydrological and energy-related analyses is centered on the use of the median. This method is preferred due to its ability to provide a clearer and more reliable representation of central tendency compared to other statistical measures, especially in the context of complex and variable hydrological data. The median offers a robust alternative to the mean, particularly in the presence of outliers or skewed distributions, which are common in hydrological datasets. By focusing on the median, it is possible to reduce the variability and uncertainties often associated with more complex or less intuitive statistical methods. This approach also helps streamline the decision-making process by providing a consistent and straightforward metric, avoiding the complications arising from the wide range of methodologies and sources found in the scientific literature. As a result, the median has proven to be a practical and effective parameter for ensuring reliable predictions in the design and operation of hydroelectric facilities, making it a key tool in the company’s analytical framework.

3.1. Theoretical references of probability and statistics

From [12], given a frequency distribution represented by y = φ(x), consider the curve:

(1) $\Phi(x) = \int_{x}^{X2} \varphi(x)dx$

Which measures the probability that a variable is between its maximum value X2 and a generic value x: that is, the probability that the variable is greater than x. This curve is called *the Duration Curve* of the x.
Another curve, closely related to the previous one, is the following:

(2) $\psi(x) = \int_{X1}^{x} \varphi(x)dx$

Which measures the probability that the variable has been found to be less than x. It is called *the Duration Complement Curve.*
From these two curves, mentioned above, it follows that:

(3) $\Phi(x) + \psi(x) = \int_{x}^{X2} \varphi(x)dx + \int_{X1}^{x} \varphi(x)dx = 1$

And, in the case of flow rates, expressing frequencies and durations in days of the year:

(4) $\Phi(q) + \psi(q) = 365$

The curves described are graphed in Figure 1.

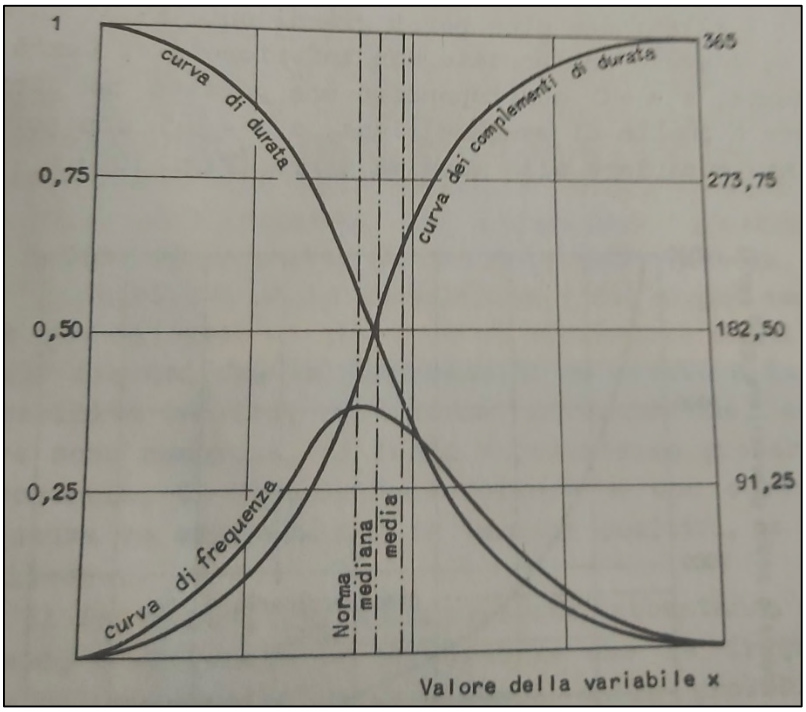

Figure 1 – frequency curve, duration curve and duration complement curve, from [12].

Figure 1 also shows the *Frequency Curve*, which, if the dataset of observations is large enough, can be referred to as the probability curve of future events, based on the assumption that: *in a series of repeated trials, the probability that an event will occur with a frequency equal to the theoretical probability tends to one (i.e. certainty) with the indefinite increase of the evidence. In short, it is a "probability within probability"*, as mentioned in [12].

3.2. Validity of the median in the statistical representation of hydrological phenomena

As can be seen in Figure 1, the two curves (duration and duration complement curve) intersect at the median of the observations. The intersection of the duration curve and the duration complement curve at the median vertical line indicates that, at that point, the probability that an event will be less than or equal to the median duration is equal to the probability that the event will be greater than the median duration.

In other words, it is a central point of distribution, because it provides a reference point that probably balances the duration of the event into two equal parts: the shorter and the longer. The assumption that the median represents a valid reference point for the processing of hydrological data is based on its ability to provide a central measure of durations without being significantly affected by extremes. Hydrological data, such as precipitation or discharges, can present an asymmetrical distribution with extremely high values that can distort the average. In these cases, the median offers a more robust and representative measure of the "common" behavior of the phenomenon under examination, without being distorted by rare but extreme events.

3.3. Reference to hypothetical median year

In relation to a duration curve, the reconstruction of the hypothetical median year involves determining the duration that separates the values above and below it equally. To make it clearer, the duration curve describes the distribution of the durations of events (e.g., rainfall or discharges) over a certain period of time, ordered from shortest to longest. The hypothetical median year represents an average event in terms of duration, which is at the point where the probability that an event will last longer or shorter than that duration is exactly the same (50% for each case). In other words, to determine the hypothetical median year on the duration curve, one must identify the duration that corresponds to the point at which the cumulative probability of having an event of shorter duration is equal to 50% of the distribution, and consequently, the probability of having an event of longer duration than the median is also 50%. This hypothetical "median year" is a useful concept for analyzing the distribution of hydrological events over an observation period, allowing to better understand the temporal distribution of event durations and to make predictions about future events with a balanced probability. In practice, it serves as a parameter to characterize a typical event in relation to its duration compared to other events in a given time interval. Comparing the hypothetical median year (determined for future forecasts) to the median of historical annual observations can be a valid approach to assess the consistency between future forecasts and historical data.

If the median of historical observations and the hypothetical median year of future predictions are similar, it means that the forecasts do not suggest significant changes from the past, at least in terms of the duration of events. If there are significant differences, it could be indicative of a change in future conditions compared to the past, such as an increase in the frequency or intensity of events.

In summary, this type of comparison can provide useful information for making decisions related to water resource management, spatial planning, and risk assessment associated with extreme events.

4. Main Advantages of Using the Median Method in Hydrological and Energy Analyses for Economic and Operational Efficiency

The use of the median as a central parameter in hydrological and energy analyses offers significant advantages in terms of both energy management and economic efficiency, particularly when applied to the assessment and management of water resources for hydroelectric power generation. One key aspect where the median plays a crucial role is in the construction and interpretation of duration curves, which are essential for understanding the distribution of hydrological events such as rainfall or river discharge over time. The duration curve, ordered from the shortest to the longest event durations, provides insights into the frequency and duration of these events, which are vital for predicting water availability and the operational performance of hydroelectric plants. By identifying the hypothetical median year on the duration curve, it can be pinpointed the duration where the probability of an event lasting shorter or longer than that point is exactly equal (50%). This method offers a practical way to assess and predict event durations, which directly impact the efficiency and capacity of energy generation from hydropower sources. The hypothetical median year concept is valuable because it represents an "average" event duration, thereby providing a more stable and representative estimate of typical hydrological events. Unlike the mean, which can be skewed by extreme values, the median is less sensitive to outliers and offers a more reliable measure of central tendency, particularly in cases where hydrological data can exhibit irregularities, such as extreme weather events or anomalous droughts and floods. This is particularly important for hydropower systems that rely on consistent water availability to generate energy. By focusing on the median, rather than more volatile or extreme measures, energy producers can better anticipate the general conditions under which their plants will operate, leading to more efficient and reliable energy generation planning.

From an energy perspective, using the median to assess the duration of events enables hydroelectric facilities to optimize their operational strategies. For example, the identification of median event durations allows hydropower plants to better align their production forecasts with expected inflow conditions. This in turn facilitates more accurate planning for energy storage, distribution, and peak load management, leading to enhanced reliability of the power grid. In particular, by understanding the temporal distribution of water flows and their corresponding durations, energy companies can predict periods of low and high generation capacity, optimizing turbine operations and reservoir management. Such predictive accuracy not only improves operational efficiency but also reduces the risk of underproduction or overproduction of energy, both of which carry significant economic costs.

Moreover, this method enables more informed decision-making regarding the allocation of water resources. By comparing the median of historical data with the hypothetical median year of future predictions, energy providers can assess how future hydrological conditions are expected to differ from the past. If the future median aligns closely with the historical median, it suggests that water availability and event durations will remain relatively stable, providing a sense of security for long-term planning and investment in infrastructure. However, if there are noticeable differences, such as an increase in event durations or a shift in their distribution, it may signal the need for adjustments in infrastructure planning or even in energy generation capacity. This could be particularly relevant in regions where climate change is influencing rainfall patterns or where significant changes in water resource availability are anticipated. The application of the median in hydrological forecasting also supports more effective risk management strategies. For hydropower plants, which are vulnerable to extreme weather events such as floods or droughts, understanding the probability and duration of such events is critical for ensuring the safety and sustainability of operations. By identifying trends in the median and using it as a benchmark for future projections, energy companies can develop better risk mitigation strategies, such as adjusting water storage levels, enhancing flood control measures, or investing in additional infrastructure to handle more extreme conditions. This not only reduces operational risks but also enhances the long-term resilience of hydropower systems to climatic variability.

In terms of economic advantages, the median-based approach is more cost-effective than relying on more complex or variable methodologies. The simplicity and robustness of the median allow energy companies to reduce the computational costs and time associated with more sophisticated statistical models, making it an attractive option for large-scale operations that require quick and efficient decision-making. Furthermore, by providing a stable and clear measure of central tendency, the median ensures that predictions and forecasts are grounded in realistic expectations, minimizing the likelihood of costly errors in energy production planning. Additionally, when used in conjunction with other hydrological and meteorological data, the median offers a solid foundation for the integration of renewable energy sources into the broader energy grid. Hydropower plants that rely on accurate and reliable water flow predictions can contribute to a more stable and predictable energy supply, reducing dependency on less reliable energy sources and enhancing grid stability. By applying median-based methods to forecast hydrological events, energy companies can improve their ability to synchronize energy generation with demand patterns, leading to reduced reliance on backup power sources and further economic savings.

In conclusion, the use of the median in hydrological and energy-related analyses offers multiple advantages, both from an operational and economic perspective. It enables more accurate, stable, and efficient energy production planning for hydroelectric facilities, supporting better risk management, resource allocation, and infrastructure investment decisions. By reducing the uncertainty associated with extreme weather events and offering a clear, reliable method for event duration prediction, the median plays a vital role in optimizing the performance and sustainability of hydropower systems. In an era of climate change and fluctuating water resource availability, this approach provides a robust and cost-effective tool for energy companies to navigate future challenges while maximizing economic returns.

5. Optimizing Hydropower and Energy Management in Italy: The Role of Median-Based Methods for Predictability and Economic Efficiency

The Italian energy market, with its diverse and evolving landscape, plays a significant role in shaping the future of hydropower and other renewable energy sources in the country. Italy, as a member of the European Union, is part of a broader energy transition that aims to reduce carbon emissions and increase the share of renewables in the energy mix. Hydropower has historically been a cornerstone of Italy's renewable energy capacity, and its importance continues to grow as part of the country's efforts to meet its energy and climate goals. The integration of statistical methods, such as the median-based approach in hydrological analysis, is crucial in optimizing the utilization of water resources, ensuring the reliability of energy generation, and contributing to the economic stability of the Italian energy market. One of the key characteristics of the Italian energy market is its reliance on renewable energy sources, with hydropower being one of the largest contributors. As of recent years, Italy has been able to generate a significant portion of its electricity from hydropower, a resource that benefits from the country's varied geography, which includes numerous rivers, lakes, and mountainous regions. Hydropower plants are strategically located throughout the country, particularly in the northern regions, where the Alps provide a steady supply of water to generate electricity. However, Italy's energy mix has been undergoing a significant transformation as part of its commitment to reducing greenhouse gas emissions. The country has seen an increase in the use of wind and solar energy, which, alongside hydropower, make up the primary sources of renewable energy. This shift toward a more diversified renewable energy landscape, while promising, also introduces new challenges in terms of grid stability and reliability, as these sources are inherently intermittent and weather-dependent.

In this context, the application of statistical methods, such as using the median to predict water availability and hydrological events, becomes increasingly important. Hydropower plants must adapt to a more variable generation profile, which can be influenced by climate change, shifts in seasonal precipitation patterns, and changing river discharge rates. For example, hydropower operators may face periods of low water availability during dry spells or droughts, which can significantly reduce energy generation potential. Conversely, heavy rainfall or snowmelt events can lead to surges in water flow, potentially resulting in excess generation and associated grid stability issues. By using the median as a predictive tool, energy producers can better manage the risks associated with both extreme events and typical hydrological variations. This method provides a clearer and more stable benchmark for estimating expected water flows and energy generation, helping to optimize the operation of hydropower plants and improve grid management.

Moreover, Italy's energy market is heavily interconnected with the broader European grid, meaning that fluctuations in energy production can have cross-border implications. The use of the median in hydrological forecasting aids in ensuring that Italian hydropower plants can maintain consistent energy production, reducing the risks of supply disruptions that could impact neighboring countries. By predicting typical event durations and water availability, the median helps ensure that hydropower plants can contribute to grid stability, even in the face of changing weather conditions or unexpected hydrological events. This is particularly important as Italy continues to integrate more renewable energy sources, such as wind and solar, into its energy mix, which are less predictable than traditional fossil fuel-based generation. The flexibility of the median-based approach in forecasting water availability allows hydropower plants to complement these intermittent sources of power and contribute to the overall reliability of the Italian electricity grid.

From an economic perspective, the benefits of applying the median in hydrological and energy analyses are multifaceted. Hydropower plants in Italy are subject to a competitive market where electricity prices fluctuate based on supply and demand. By accurately forecasting water availability and adjusting energy generation accordingly, operators can optimize their operations to take advantage of favorable market conditions, such as high electricity prices during periods of peak demand. For instance, during a dry year or a period of low rainfall, hydropower operators can adjust their expectations and reserve water for high-demand periods, ensuring they can still provide power when prices are higher. Conversely, during wetter periods when water flow is more abundant, they can maximize generation and take advantage of lower market prices, contributing to overall cost-effectiveness. The use of median-based forecasting methods also supports more efficient water resource management, which is essential in a country like Italy where water availability can be highly variable. By focusing on the median, which reflects typical conditions, operators can make more informed decisions about how much water to store or release from reservoirs, avoiding both the risk of flooding during peak flow events and the risk of energy shortages during low-flow periods. This balanced approach to water management not only supports operational efficiency but also contributes to the long-term sustainability of hydropower as a reliable energy source in Italy.

Furthermore, the financial benefits of using the median extend beyond operational optimization. The application of robust and reliable statistical methods helps to mitigate risks associated with hydrological uncertainties, which are crucial for attracting investment in hydropower infrastructure. Investors and stakeholders in Italy’s energy sector are increasingly looking for stable, predictable returns on investments, especially in the context of renewable energy. By demonstrating the ability to manage water resources effectively and generate consistent power, hydropower operators can enhance investor confidence and secure funding for new projects or the modernization of existing plants. The use of statistical tools like the median strengthens the credibility of investment prospects by offering a clear, scientifically grounded method for predicting future performance.

In summary, the Italian energy market, with its growing emphasis on renewable energy, stands to benefit significantly from the use of median-based methods in hydrological analysis. These methods improve the predictability and reliability of hydropower generation, ensuring that Italy’s energy needs can be met sustainably and economically. By incorporating such methods into energy production strategies, hydropower operators in Italy can contribute to grid stability, optimize energy production in line with market conditions, and enhance water resource management. As Italy continues to transition toward a cleaner and more diversified energy mix, the median offers a valuable tool for addressing the inherent uncertainties of renewable energy generation and ensuring the long-term viability of hydropower as a key component of the country’s energy strategy.

6. Conclusion

In conclusion, the analysis of the hypothetical median year and the intersection of the duration curves and duration complement curves at the median allows to obtain a complete picture of the temporal behavior of both historical and future hydrological events. The hypothetical median year represents a value that, based on future forecasts, equally separates events into shorter and longer durations, serving as a reference point for a balanced forecast. This value can be compared to the median of historical observations, which represents the point at which the distribution of past events balances out, separating shorter events from longer ones.

The intersection of the duration curves and duration compliments, at the median, indicates the point at which the cumulative probabilities of lower and longer durations are equal, marking the central value in the distribution. When comparing the median of historical observations with the hypothetical median year, a useful assessment can be made of the consistency between the past and future projections.

If the two medians are similar, it means that no significant change is expected from the past. Otherwise, if there are notable differences, there may be an indication of changes in hydrological or climatic conditions, such as increased intensity or frequency of extreme events. This understanding helps make informed decisions about water resource management, spatial planning, and mitigation of risks associated with future hydrological events. Additionally, from an energy perspective, the application of such forecasting techniques in the context of Italy's energy market enables better hydropower production planning, ensuring grid stability and optimizing energy generation in line with variable water availability, which is crucial for maintaining economic efficiency and securing a reliable energy supply in a renewable-driven market.

7. References

[1]. De Michele C and Salvadori G (2003) A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. Journal of Geophysical Research 108(D2): 4067 (doi:10.1029/2002JD002534).

[2]. Berryman D, Bobee B, Cluis D, and Haemmerli J (1988) Nonparametric tests for trend detection in water quality time series. AWRA Water Resources Bulletin 24(3): 545--556.

[3]. Cluis D, Langlois C, van Coillie R, and Laberge C (1989) Development of a software package for trend detection in temporal series: Application to water and industrial effluent quality data for the St. Lawrence river. Environmental Monitoring and Assessment 12: 429--441.

[4]. Helsel DR and Hirsch RM (1992) Statistical Methods in Water Research. Amsterdam: Elsevier.

[5]. Salas JD (1993) Analysis and modeling of hydrologic time series. In: Maidment DR (ed.) Handbook of Hydrology. New York: Mc-Graw Hill.

[6]. Fanta B, Zaake BT, and Kachroo RK (2001) A study of variability of annual river flow of the southern African region. Hydrological Sciences Journal 46(4): 513--524.

[7]. Yue S, Pilon P, and Cavadias G (2002) Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. Journal of Hydrology 259: 254--271.

[8]. Fuller W (1914) Flood flows. Transactions of the American Society of Civil Engineers 77: 564--617.

[9]. Gumbel EJ (1958) Statistics of Extremes. New York: Columbia University Press.

[10]. Klemesˇ V (1993) Probability of extreme hydrometeorological events – a different approach. In: Kundzewicz Z (ed.) Extreme hydrological events: Precipitation, Floods and Droughts, vol. 213, pp. 167--176. Wallingford: IAHS.

[11]. Salvadori G and De Michele C (2007) On the use of copulas in hydrology: Theory and practice. Journal of Hydrologics Engineering 12(4): 369--380 (doi:10.1061/(ASCE)1084-0699(2007)12:4 (369)).

[12]. "Impianti idroelettrici", Vol.1, G. Evangelisti, Patron, 1982.