Applications of extreme value theory for dynamical systems to the analysis of blood pressure data.

Davide Faranda

30

Applications of Extreme Value Theory in Dynamical Systems for the Analysis of Blood Pressure Data

Davide Faranda

30.1 Introduction

Over the last few decades, several advancements in experimental techniques have allowed for the measurement of health parameters at a high sampling frequency [3,5]. Although a frequency analysis of the data often reveals major anomalies and helps in the diagnostic process, a great part of the information is hidden in fine properties of the measured time series. Among these properties, extreme fluctuations of health parameters may trigger irreversible processes and result in acute crisis. In the ambit of cardiovascular disease, blood pressure fluctuations may trigger acute hypotensive (hypertensive) episodes and, in some cases, cardiac crisis. For a series of independent and identically distributed (iid) variables, a traditional extreme value analysis straightforwardly gives the probability of observing extremely low (or high) fluctuations of health parameters [13]. However, blood pressure data have internal correlations originating from the quasiperiodic biological processes responsible for blood circulation. Therefore, in order to provide effective warnings against cardiac crisis, the traditional techniques must be accompanied by methods that preserve the dynamical information contained in the data.

In previous studies, the risk of observing acute hypotensive or hypertensive episodes has been assessed either by analyzing blood pressure data averaged over 1 minute [16], or by using neural network multimodels [11], or via spectral techniques [2]. All these techniques usually rely on the identification of a single-threshold pressure value defined by counting the exceedances (hypertensive episodes) or the deficits (hypotensive episodes) with respect to such a threshold. Statistics are always computed under the nongenuine assumption that pressure data are iid. However, when this assumption is satisfied, these methods would only provide information on the tails of the distribution, without inspecting the bulk statistics. This would prevent from providing a global map of the status of the patient, providing the nursing staff with

the detailed conditions of the patient. Moreover, they do not associate the probability of extreme pressure events to a time scale of medical interest, e.g., length of a medical treatment or of an operation.

In this chapter we show that, by combining the celebrated theory of Poincaré recurrences [9] with extreme value statistics, one can devise an efficient algorithm to measure the range of expected fluctuations of systolic and diastolic arterial pressures. Such a range can be used to assess the clinical state of a patient and therefore the condition of the patient. The combination of the extreme value statistics with new results from dynamical systems theory allow for a proper treatment of correlations in the data, without the assumption that data are iid. The chapter is structured as follows: Section 30.2 gives an overview on the method and describe the link between recurrences of blood pressure values and rare events. In Section 30.3 we present the multiparameter intelligent monitoring in intensive care (MIMIC) database and describe some general properties of the blood pressure time series. Section 30.4 is dedicated to the data analysis and the definition of an index that directly measure the probability of observing the likelihood for recurrences of pressure values. We conclude by discussing our findings.

30.2 The Method

Here, we briefly give some theoretical elements of the method of recurrences. For a time series $p(t)$ we start by defining the common approaches in the definition of a recurrence of a certain value p^*. Such a description is accompanied by direct visualization of the different methods in Figure 30.1.

- The *exact recurrence* of the value p^* occurs if we have $p(t) = p^*$. This definition is very restrictive as it requires to observe identical values of p^*. In the upper panel of Figure 30.1 exact recurrences are found on the intersections between the orbits and the solid horizontal line.
- The *spatial approach* consists of choosing an interval Δp such that $p(t)$ is a recurrence of $p(t^*)$ if $p^* - \Delta p < p(t) < p^* + \Delta p$. In Figure 30.1, it corresponds to considering all the values falling in the solid rectangle. The spatial approach is equivalent to the so-called peak-over-threshold technique for extremes introduced by Pickands [17]. Extracting the recurrences this way correspond to sampling the statistics of the distances $\mathrm{dist}(p(t), p^*) < \Delta p$, with Δp acting as a threshold value. This statistics will depend on the choice of Δp and instead it will not be sensible on the dynamics. A random reshuffle of the time series will reproduce exactly the same threshold statistics, while destroying all the correlations in the dataset.
- The *time window approach* does not imply the selection of a threshold Δp but rather the introduction of a time scale τ. The recurrence is defined as the closest value of p^* in the interval τ, formally $\min(\mathrm{dist}(p(t), p^*))$ for $t^* < t < t^* + \tau$. We construct the statistics of such recurrences by extracting n values in all intervals of length τ. These points are marked by the circles in the lower panel of Figure 30.1.

We stick on to the latter approach because it preserves the dynamical information contained in the data, i.e., a reshuffle of the data does not produce the same asymptotic distribution as it happens for the Pickands method. In the limit of $n, \tau \to \infty$, a series of minima

$$X_i = \min(g(\mathrm{dist}(p(t), p^*))) \quad t_i < t < t_i + \tau$$

with $g(\cdot)$ a generic observable function, will obey the generalized extreme value (GEV) distribution:

$$F_G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}. \tag{30.1}$$

A GEV holds for $1 + \xi(x - \mu)/\sigma > 0$, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter, and ξ the shape parameter that discriminates the type of tail behavior: Gumbel law for bulk statistics with

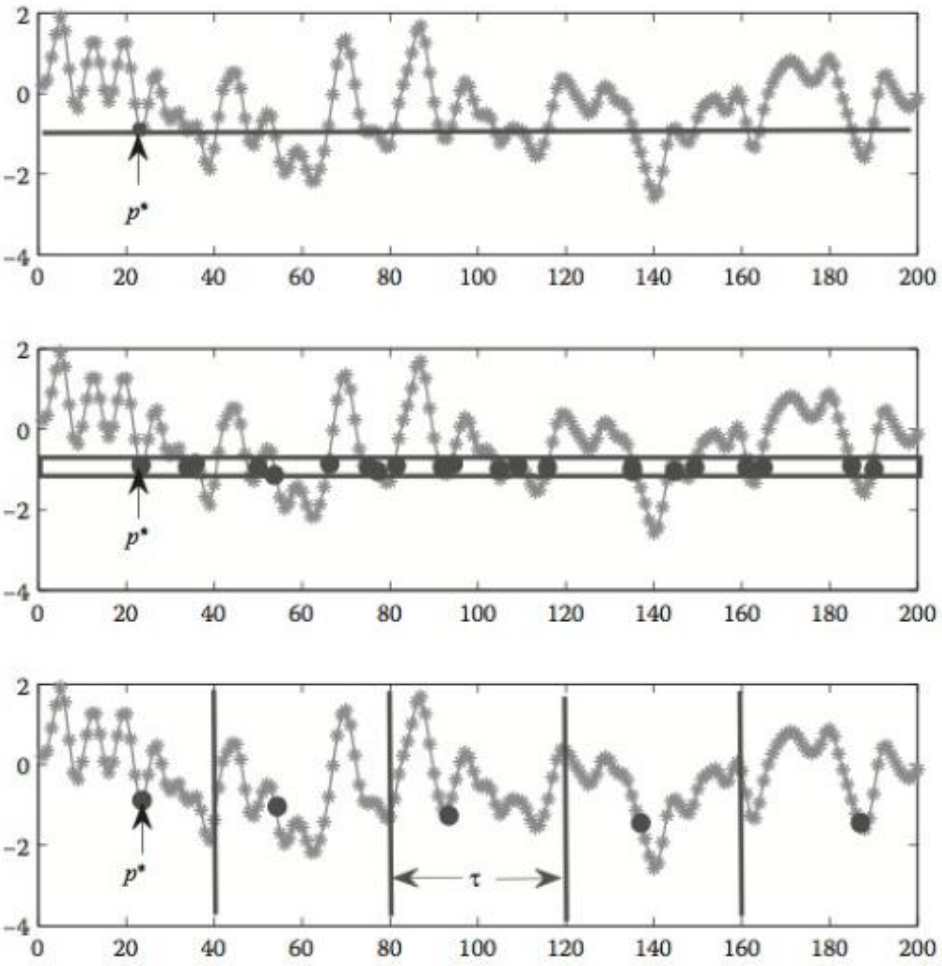

FIGURE 30.1 Example of three different definitions of recurrences of the value p^* of a time series. Upper panel: exact recurrences occur if $p(t) = p^*$ (points on the solid line). Central panel: recurrence of p^* occurs each time $p^* - \Delta p < p(t) < p^* + \Delta p$ (points within the solid rectangle). Lower panel: recurrences are sampled over a time interval τ as the closest values to p^* are measured in each interval.

exponential tails ($\xi = 0$), Fréchet laws for unbounded tails ($\xi > 0$), and Weibull laws for tails bounded from above ($\xi < 0$). As shown by in References 6 through 10, the convergence is given by a combination of the exponential recurrences statistics for chaotic systems and the functional form of the observable function $g(\cdot)$. In particular, letting $g(\cdot) = -\log(\cdot)$, the asymptotic GEV is always a Gumbel law with the shape parameter $\xi = 0$ [9]. The use of the logarithmic function to measure the distances is not arbitrary and it has been justified in several papers as it allows for a correct sampling of the so-called short returns [1,20]. This result is used to answer the following question:

Given a time series representing a chaotic dynamics, and an observation p^*, what is the typical time scale τ such that this observation is recurring in the dynamics?

We can answer by measuring the likelihood of a fit of the X_i's to the Gumbel model, provided that we have extracted the data as described above:

- If the fit succeeds, it means that we are likely to observe a recurrence of p^* in τ. We can repeat the experiment for shorter bin lengths and find the smallest τ such that the fit converges. This defines the shortest convergent recurrence time.
- If the fit fails, the recurrence of p^* in τ is unlikely. We can repeat the experiment by increasing the value of τ until the fit succeeds to find the shortest recurrence time.

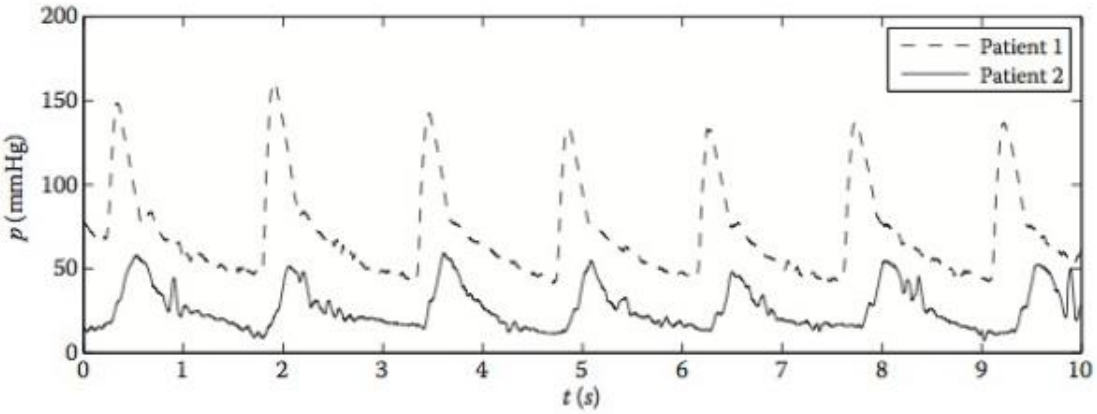

FIGURE 30.2 An example of time series of systemic arterial pressure (dashed line) and pulmonary arterial pressure (solid line) for a patient with unstable post infarction angina.

For the applicability of this method on blood pressure data, first, we have to check the chaotic behaviors of the time series, and then assess the typical time scales involved in the process. For this purpose, in the next section, we perform a first analysis of the blood pressure data, which will clarify these two important issues.

30.3 The Blood Pressure Data

The data have been extracted by the MIMIC II database [19] that includes, for each patient, a time series of systemic and pulmonary arterial pressures sampled at 300 Hz. Data have been collected over a period of about 1 h. The general characteristics of the time series analyzed are presented in Figure 30.2. The series refers to patient 28 of the MIMIC database (*mgh028*), recovered at the Massachusetts General Hospital (MGH) for unstable postinfarction angina. We isolate only 10 s of the time series corresponding roughly to $3 \cdot 10^4$ observations. It is apparent that data are not iid: the periodic signature of the heartbeat is evident, although deviations from the periodic behavior both in time and amplitude can be observed. On longer time scales, ambient conditions, human interactions, and/or pharmacological treatments trigger a chaotic behavior that legitimates the assumptions made in the previous sections. Such chaotic behavior of systemic pressure data has been reported by several authors [21,22]. This first analysis also provides a minimal time scale for the recurrence technique: we will need to consider a period τ longer than the heartbeat oscillation in order to properly sample recurrences of a chaotic system. In the next section, we will start the recurrence analysis by analyzing the effects of τ on the statistics.

30.4 Analysis

Our goal is to provide an accurate information on the status of patients by applying the recurrence techniques described in the previous sections. The final product of such an analysis will be a map of health parameters (here systemic and pulmonary arterial pressures) that inform the nursing staff on the status of the patient and whether he experienced anomalous pressure values with respect to a certain time scale of interest τ. We recall the procedure applied on the data in an algorithmic way:

1. We divide the full time series (the length is typically 1 h) in n intervals of length τ.
2. We fix a reference pressure value p^*.
3. In each interval, we take the closest value of $p(t)$ to the chosen p^*, obtaining n values $p_1, p_2, \ldots, p_n$ as shown in Figure 30.1.
4. We construct a series of logarithmic returns for such values, namely $X_i = -\log(\text{dist}(p_i, p^*))$.
5. We fit the X_i to the GEV model checking the goodness of the Gumbel model.

We repeat the analysis over and over changing the reference points, until all the p^* between $min(p)$ and $max(p)$ have been selected. As mentioned before, if there are enough recurrences of a certain pressure value in τ, the fit will succeed; otherwise it will fail. As a measure of the goodness of the Gumbel model, we use a suitably renormalized likelihood function. This provides a measure of the probability that a patient will experience a pressure value p_{ART} in a time interval τ. The distance between the Gumbel law and the histograms will be quantified by the quantity δ defined as

$$\delta(p_{ART}) = \frac{L(p_{ART})}{L_{max} - L_{min}}$$

Here, $L_{p_{ART}}$ is the negative likelihood function of observing a Gumbel distribution for the recurrences of the value $p^* = p_{ART}$. Since L attains infinity for a *perfect* fit and minus infinity for a completely unreliable fit, then δ takes values in the range $(-1, 1)$, so that

- $\delta(p_{ART}) > 0$ indicates how likely is the recurrence of the value p^* every τ seconds.
- $\delta(p_{ART}) < 0$ indicates how unlikely the value p^* recurs in a τ time interval.

First, we verify the consistency of our method with respect to the usual pressure measurements. The maxima of δ should correspond with the measured systolic and diastolic pressures, respectively p_s and p_d, and they should not sensibly change with τ. This is what is shown in Figure 30.3 for the series of patient *mgh002* and for three different values of τ. This patient was admitted to the hospital for resection and grafting of abdominal aortic aneurysm and his measured systemic pressure parameters were $p_d =$ 60 mmHg and $p_s = 130$ mmHg. Effectively, from Figure 30.3 we observe two maxima of δ corresponding to p_s and p_d. However, although the value of p_d is consistent with the one measured at the hospital, the estimate for the systolic pressure is 150 mmHg higher than the one estimated using classical techniques. This is explainable by looking at the shape of the distribution of δ around the p_s value that is skewed toward values smaller than p_s. We recover the value of p_s measured at the hospital by taking a conditional mean of p_{ART} restricted to values higher than 80 mmHg.

These first considerations already point to the problem of measuring only average values of pressure, rather than fully characterizing the probability distributions as we attempt to do here. Moreover, the analysis of δ clearly points to a large range of fluctuations around the p_s-detected value and suggests that this patient could experience hypertensive episodes, underestimated with the classical measurements. We also observe that the parameters slightly depend on τ and that for the shortest time interval $\tau = 1$ s curves of δ are not smooth. This effect can also be detected for other patients and it is explainable with the general argument pointed out in the previous section, i.e., τ must be larger than the heartbeat period in order to get a good convergence for δ. For this reason, in the remaining of this chapter, we will fix $\tau = 4$ s to avoid such problems. We remark that the parameter τ can be varied to correspond to a specific time of interest for medical treatments.

To show the capability of the method, we analyzed several other patients presenting different clinical histories. In Figure 30.4 we present the quantity δ for three particular subjects: patient 1—*mgh002* considered in the previous analysis (left panels); patient 2—*mgh001* presenting a carotid endarterectomy (central panels); and patient 3—*mgh028* showing an unstable postinfarction angina (right panels). In all the cases, the two maxima corresponding to the systolic and diastolic pressures are recognizable. However, their location and the range of fluctuations associated with each of them strongly varies. For patient 2, the hemodynamic data recorded in the database are $p_s = 140$ and $p_d = 50$ mmHg, in agreement with our observations. The analysis also points to a moderate hypertensive risk. For patient 3, we observe a slight deviation from the p_s value reported in the database ($p_s = 130$ and $p_d = 50$ mmHg) again due to the skewed nature of the distribution of the systolic blood pressure already observed in References 14 and 18.

The general medical case of a patient is not given by a single hemodynamic parameter but rather by a combination of them. One of the advantages of the recurrence technique described in this chapter is that it is trivially expandable to multiple hemodynamic parameters. As an example, we consider both the arterious p_{ART} and the pulmonary pressure p_{AP}, and define the joint probability of observing a Gumbel

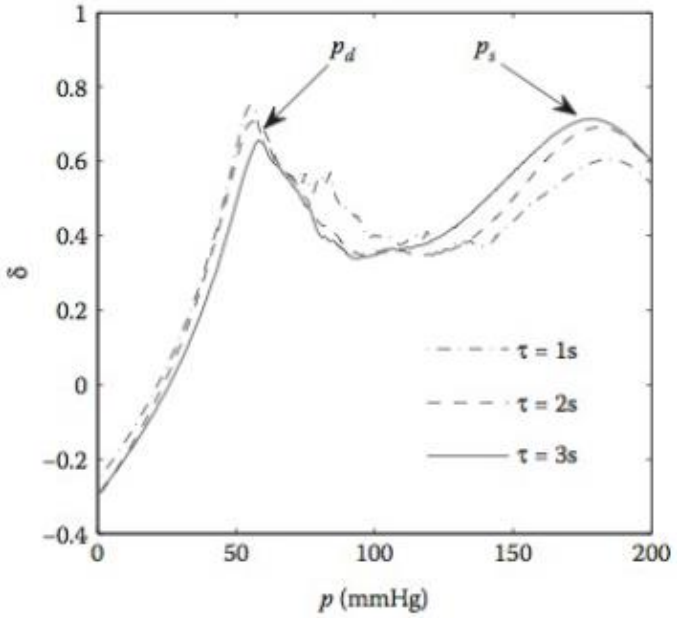

FIGURE 30.3 Normalized likelihood of the Gumbel distribution δ for three different values of $\tau = 1, 2, 3$ s. Maxima indicate the location of the estimated systolic pressure p_s and diastolic pressure p_d.

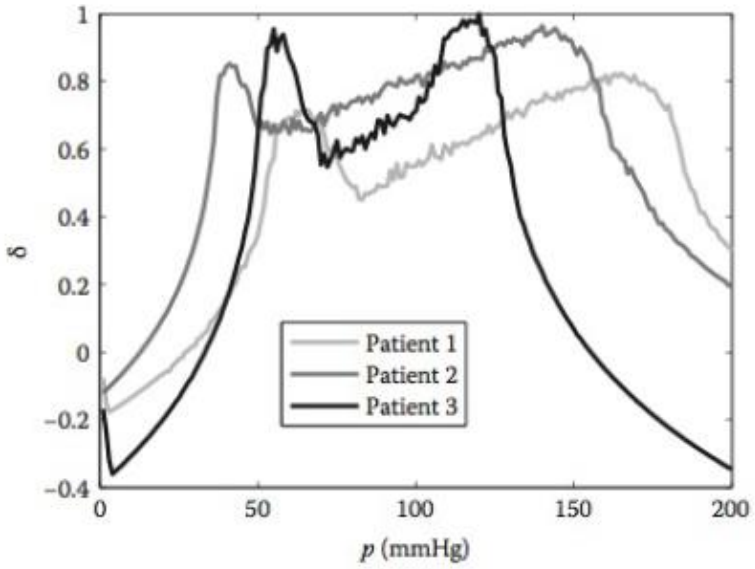

FIGURE 30.4 Normalized likelihood of the Gumbel distribution δ of the systemic arterial pressure data p_{ART} for three different patients: patient 1 admitted to the hospital for resection and grafting of an abdominal aortic aneurysm, patient 2 for carotid endarterectomy, and patient 3 for unstable postinfarction angina. Time interval is $\tau = 4$ s.

distribution for the recurrences of a specific couple of values $p^* = (p_{ART}, p_{AP})$. The structure of the function $\delta(p_{ART}, p_{AP})$ will therefore give an immediate information on the joint risk of experiencing pulmonary and systemic hypertension (or hypotension). For the three patients previously described, the results are reported in Figure 30.5. The left panel refers to patient 1: light gray indicate that the probability of observing a couple of values (p_{ART}, p_{AP}) is high, dark gray that this combination is unlikely. For this patient, there exists a moderate risk of experiencing a combination of pulmonary and systemic hypertensive events. The analysis of patient 2 (central panels) still points to a risk of moderate systemic hypertension. The analysis of pulmonary pressure reveals that this risk is also associated with pulmonary hypertension: the $\delta(p_{ART}, p_{AP})$

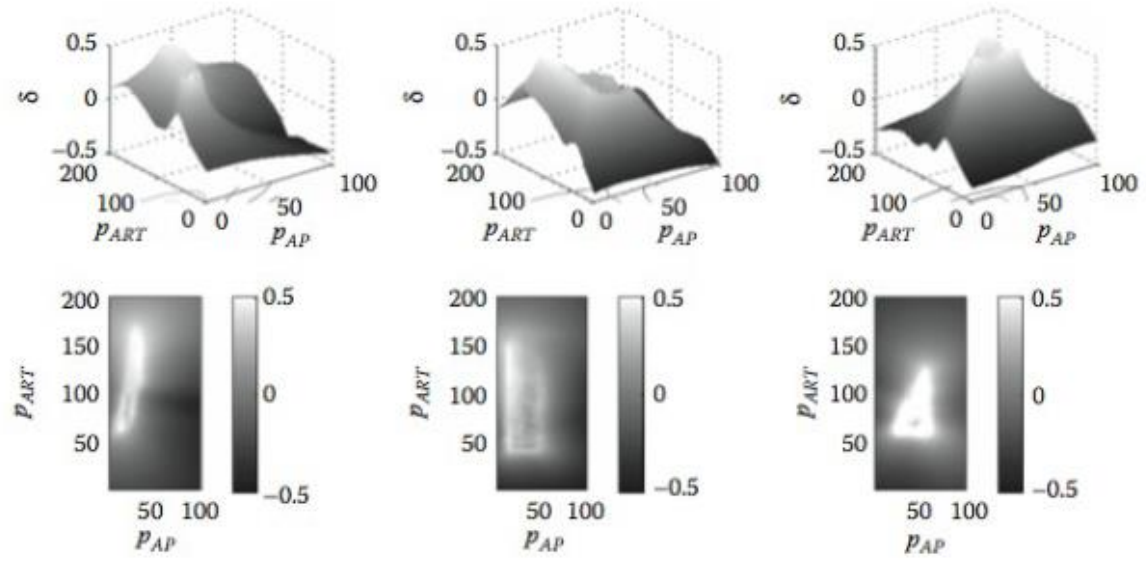

FIGURE 30.5 Normalized likelihood of the Gumbel distribution δ for three different patients obtained by combining systemic arterial p_{ART} and pulmonary arterial p_{AP} pressure data. Time interval is $\tau = 4$ s. Left panels: patient with resection and grafting of an abdominal aortic aneurysm. Central panels: patient with carotid endarterectomy. Right panels: patient with unstable postinfarction angina. Positive values of δ mean that we are very likely to observe the corresponding values of pressure, negative values that they are unlikely.

profile consists of a sort of ring of probable values, i.e., the bimodal structure observed for the pressure data also repeats for the pulmonary pressure. Finally, $\delta(p_{ART}, p_{AP})$ for patient 3 shows a more complicated pattern associated with a higher risk of hypertensive pulmonary episodes with respect to blood arterial episodes. This is also reported in the logbook of the MGH database.

30.5 Final Remarks

The main achievement of this chapter is to provide a method for the detection of recurrences of pressure values. The recurrence technique shows that the concept of systolic and diastolic blood pressures commonly used in the medical care units can be insufficient for assessing the risk of hypotensive and hypertensive episodes. The main drawback of pure statistical techniques is the misleading assumption that the data are iid, a problem that we overcome by using an approach based on recurrences. Here, we have introduced a parameter δ directly linked to the likelihood of observing a Gumbel statistics expected on a theoretical basis for the recurrences of pressure values. Positive values of δ are associated with higher probabilities and negative values are associated with lower ones. In other words, the profiles of δ provide the following information: for each pair of values p_{ART}, P_{AP}, a positive δ means that, if the conditions of the patient stay stationary, it is likely to observe such values again. This allows to take actions if there is a risk for acute episodes and therefore they provide a useful instrument for operators of medical care units. With respect to the common methods, based on computing the histograms and the probability distribution of pressure values, the study of recurrences provides the following advantages: (i) the probability of observing a certain pressure value is given with respect to a certain time scale τ and condensed in an index with values of $[-1 +1]$. This means that one can plan medical treatments, operations, or examinations that last for a specific time knowing the risks associated with the patient. (ii) It allows for combining several variables giving the joint probability of observing a combination of health parameters in the time scale of interest τ. (iii) It provides a precise information on the quantity of data needed to assess the Gumbel statistics. This prevents one from overestimating or underestimating the recurrence times. (iv) The technique is easy to implement—a self-written MATLAB code has been used to perform the analysis—and can run

in real time. Profiles of δ can be updated with new data, e.g., every minute, giving a map for the changing risk for each patient. This technique can be easily extended to any framework where chaotic time series are available even if they show fractal or multifractal properties and/or power law behavior (see, e.g., [4,12]) as theoretically described in Reference 15. In these cases, traditional statistical approaches are more likely to return biased estimations as they implicitly assume continuous support for the time series as well as independence among the data.

References

1. M Abadi and S Vaienti. Large deviations for short recurrence. *Discrete and Continuous Dynamical Systems—Series A*, 21(3):729–747, 2008.
2. S Axelrod, M Lishner, O Oz, J Bernheim, and M Ravid. Spectral analysis of fluctuations in heart rate: An objective evaluation of autonomic nervous control in chronic renal failure. *Nephron*, 45(3):202–206, 1987.
3. G Beevers, GYH Lip, and E O'Brien. Blood pressure measurement. *British Medical Journal*, 322(7293):1043–1047, 2001.
4. JT Bigger, RC Steinman, LM Rolnitzky, JL Fleiss, P Albrecht, and RJ Cohen. Power law behavior of rr-interval variability in healthy middle-aged persons, patients with recent acute myocardial infarction, and patients with heart transplants. *Circulation*, 93(12):2142–2151, 1996.
5. AV Chobanian, GL Bakris, HR Black, WC Cushman, LA Green, JL Izzo, DW Jones, BJ Materson, S Oparil, and JT Wright. Seventh report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. *Hypertension*, 42(6):1206–1252, 2003.
6. D Faranda, X Leoncini, and S Vaienti. Mixing properties in the advection of passive tracers via recurrences and extreme value theory. *Physical Review E*, 89(5):052901, 2014.
7. D Faranda and S Vaienti. A recurrence-based technique for detecting genuine extremes in instrumental temperature records. *Geophysical Research Letters*, 40(21):5782–5786, 2013.
8. D Faranda and S Vaienti. Extreme value laws for dynamical systems under observational noise. *Physica D: Nonlinear Phenomena*, 280:86–94, 2014.
9. ACM Freitas, JM Freitas, and M Todd. Hitting time statistics and extreme value theory. *Probability Theory Related Fields*, 147(3):675–710, 2010.
10. ACM Freitas, JM Freitas, and M Todd. Extreme value laws in dynamical systems for non-smooth observations. *Journal of Statistics Physics*, 142:108–126, 2011.
11. J Henriques and TR Rocha. Prediction of acute hypotensive episodes using neural network multi-models. In *Computers in Cardiology*, p. 549–552. IEEE, Park City, UT, 2009.
12. HV Huikuri, TH Mäkikallio, CK Peng, AL Goldberger, U Hintze, and M Møller. Fractal correlation properties of rr interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction. *Circulation*, 101(1):47–53, 2000.
13. MR Leadbetter. Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete*, 65(2):291–306, 1983.
14. Y Li, JG Wang, E Dolan, PJ Gao, HF Guo, T Nawrot, AV Stanton, DL Zhu, E O'Brien, and JA Staessen. Ambulatory arterial stiffness index derived from 24-hour ambulatory blood pressure monitoring. *Hypertension*, 47(3):359–364, 2006.
15. V Lucarini, D Faranda, G Turchetti, and S Vaienti. Extreme value theory for singular measures. *Chaos*, 22(2):023135, 2012.
16. GB Moody and LH Lehman. Predicting acute hypotensive episodes: The 10th annual physionet/computers in cardiology challenge. In *Computers in Cardiology*, p. 541–544. IEEE, Park City, UT, 2009.
17. J Pickands III. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

18. SC Robinson and M Brucer. Range of normal blood pressure: A statistical and clinical study of 11,383 persons. *Archives of Internal Medicine*, 64(3):409–444, 1939.
19. M Saeed, M Villarroel, AT Reisner, G Clifford, LW Lehman, G Moody, T Heldt, TH Kyaw, B Moody, and RG Mark. Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access intensive care unit database. *Critical Care Medicine*, 39(5):952, 2011.
20. B Saussol. An introduction to quantitative Poincaré recurrence in dynamical systems. *Reviews in Mathematical Physics*, 21(08):949–979, 2009.
21. VK Yeragani, M Mallavarapu, RKA Radhakrishna, M Tancer, and T Uhde. Linear and nonlinear measures of blood pressure variability: Increased chaos of blood pressure time series in patients with panic disorder. *Depression and Anxiety*, 19(2):85–95, 2004.
22. U Zwiener, D Hoyer, R Bauer, B Lüthke, B Walter, K Schmidt, S Hallmeyer, B Kratzsch, and M Eiselt. Deterministic–chaotic and periodic properties of heart rate and arterial pressure fluctuations and their mediation in piglets. *Cardiovascular Research*, 31(3):455–465, 1996.