

Assessing the Proximity of Time Evolutions through Dynamic Time Warping

Davide Fabozzi ¹ and Thierry Van Cutsem ²

¹ D. Fabozzi (davide.fabozzi@ulg.ac.be) : Department of Electrical Engineering and Computer Science (Montefiore Institute), University of Liège, Sart Tilman B37, B-4000 Liège, Belgium

² T. Van Cutsem (t.vancutsem@ulg.ac.be) : Fund for Scientific Research (FNRS) and Department of Electrical Engineering and Computer Science of the University of Liège, B37, B-4000 Liège, Belgium

Abstract. Comparing time-series is a frequent task in many scientific fields. In power systems, in particular, it may be of interest to compare the outputs of a simplified and a detailed model, or to validate the output of a model with respect to a measured time response. The classical Euclidean distance, involving pairs of points of the two data series aligned in time, is not suited to the practical time evolutions met in power systems, which often involve variable time delays and jumps at discrete times. In this paper, an alternative measure of proximity, stemming from other scientific fields, is proposed for power system applications. It consists in warping the time axis to guarantee the best match between the two time-series, i.e. it maps points on two curves that are not aligned in time so as to minimize the sum of squared differences of their ordinates. Modifications and adaptations of the classical algorithm to better fit power system problems are discussed. The method is illustrated through three representative curve comparison problems. A multi-dimensional extension allowing system-wide measures of similarity is also proposed.

Keywords. Time-series, curve matching, proximity measure, Euclidean distance, dynamic time warping, dynamic programming, system-wide validation.

1 Introduction

Comparison of time-series occurs in numerous scientific disciplines. The problem is to assess how close a dynamic evolution characterized by a time-series $\mathbf{q} = [q_1, q_2, \dots, q_n]$ is to a reference characterized by a time-series $\mathbf{r} = [r_1, r_2, \dots, r_n]$. The former is referred to as the query and the latter as the reference time-series.

In the power system literature, the accuracy of query with respect to reference is often depicted by qualitative, and admittedly subjective, evaluation terms such as “we see a good overall match between \mathbf{q}_1 and \mathbf{r} ”, “ \mathbf{q}_2 and \mathbf{r} have nearly the same degree of accuracy” or “ \mathbf{q}_3 is almost undistinguishable from \mathbf{r} ”, etc. The objective of this paper is to promote the use of a less subjective measure of proximity.

Assuming that both time-series span the same time interval and are uniformly sampled at the same n time instants, a simple measure of proximity is the Euclidean distance between \mathbf{q} and \mathbf{r} defined as:

$$E(\mathbf{q}, \mathbf{r}) = \sqrt{\sum_{i=1}^n (q_i - r_i)^2} \quad (1)$$

In many cases, this distance is not an effective measure of similarity. For instance, for the two time-series depicted in Fig. 1, a mere visual inspection shows that the two curves are merely shifted in time, which is not significant in usual power system applications. While both curves should be considered quite close to each other, the Euclidean distance does not recognize the pattern similarity and takes on a large value.

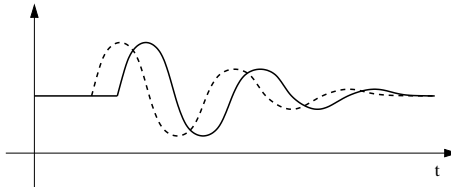


Figure 1: Two shifted damped sinusoids

More complex validation metrics that can overcome this problem are used in the field of computational mechanics: the metrics reviewed in [1] give a measure in terms of magnitude and phase differences. This would correctly identify the similarity between the curves in Fig. 1. However, the measure of magnitude and phase differences correctly identifies a time shift only if (i) the latter is constant throughout the whole time interval, and (ii) the initial and final values coincide. Unfortunately, both assumptions are seldom valid for power system time responses, where: (i) discrete events can take place at slightly different time instants, (ii) the time-shift may vary in time, and (iii) the system may settle to a different state or even collapse after a disturbance. A typical voltage

collapse case, where neither Euclidean distance nor magnitude and phase difference measures would perform well, will be assessed later on in Section. 5.1. Thus, the above quoted measures of proximity are not suited to compare time-series involving events and phenomena typical of power system dynamics.

An interesting alternative is Dynamic Time Warping (DTW), an algorithm originally developed some decades ago for speech recognition [2, 3] and now used in various fields such as medicine [4], data mining [5], signature verification [6], and others.

DTW warps the time axis to guarantee the best match between both time-series. More precisely, it maps the points of both curves so that the sum of their squared ordinate differences is minimum, under some constraints. This is equivalent to determining the deformation of the time axes which brings the two time-series as close as possible in the Euclidean-norm sense. The minimization is performed with the help of an efficient dynamic programming algorithm.

The so obtained, more rigorous measure of proximity has several potential applications in power system dynamics. For instance, in the context of simplified simulations, it can be used to evaluate the relative accuracy of one simulation with respect to another. In the context of model validation, DTW yields an objective measure of proximity between a simulated response and a measured reference. It can be also used to classify power quality events [7] or to detect instability modes [8].

The paper is organized as follows. Section 2 describes the principle and determination of the DTW distance. In Section 3 some algorithmic improvements and adaptations to typical power system evolutions are proposed, while additional aspects are discussed in Section 4. Examples of application to solver and model validation are provided in Section 5. The paper ends with the Conclusion in Section 6.

2 Principle and determination of DTW distance

2.1 Definition of DTW distance

For the sake of simplicity, the query and reference time-series are supposed to cover the same time span, have the same number n of points, and be uniformly sampled. Note that the method can be adapted to the case of non-uniform samplings or different time spans. Alternatively, the time-series could be cropped and uniformly re-sampled before applying the method. Further discussions on sampling can be found in Section 4.

Let us define the squared distance between one point of \mathbf{q} and one point of \mathbf{r} as :

$$\ell_{(i,j)} = (q_i - r_j)^2 \quad i, j = 1, \dots, n \quad (2)$$

The Euclidean distance in between the two time-series is given by:

$$E(\mathbf{q}, \mathbf{r}) = \sqrt{\sum_{i=1}^n \ell_{(i,i)}} \quad (3)$$

Relaxing the constraint that the associated points of \mathbf{q} and \mathbf{r} relate to the same time instants, we define a warping path as an N -dimensional array of pairs

$$\mathbf{w} = [w_1 \ w_2 \ \dots \ w_N] = [(i_1, j_1) \ (i_2, j_2) \ \dots \ (i_N, j_N)] \quad (4)$$

where the first element of each pair is an index referring to a point of \mathbf{q} and the second element to a point of \mathbf{r} . In addition, a warping path obeys the following three constraints:

1. *Boundary condition.* The path starts at the initial point of both time-series, i.e. $w_1 = (1, 1)$, and ends at the final point of both time-series, i.e. $w_N = (n, n)$. In Section 3 the requirement on the final point will be somewhat relaxed to cope with evolutions not ending at the same time.
2. *Continuity.* Given $w_k = (i, j)$ and $w_{k+1} = (i', j')$, the following inequalities hold true: $i' \leq i + 1$ and $j' \leq j + 1$. Thus, the path does not “skip” any point of either time-series.
3. *Monotonicity.* Given $w_k = (i, j)$ and $w_{k+1} = (i', j')$, the following inequalities hold true: $i' \geq i$ and $j' \geq j$ and $(i' - i) + (j' - j) \geq 1$. Thus, the path neither “goes back in time” nor “stalls”.

For a given warping path \mathbf{w} , one can define the distance D between \mathbf{q} and \mathbf{r} along that path as:

$$D(\mathbf{q}, \mathbf{r}, \mathbf{w}) = \sqrt{\sum_{k=1}^N \ell_{w_k}} \quad (5)$$

where ℓ_{w_k} is the quantity defined by (2) for the pair (i, j) present in w_k , the k -th component of \mathbf{w} .

As an illustrative example, consider the two time evolutions shown in Fig. 2.c. A warping path is shown with dotted lines. This path can be better visualized on the local distance matrix \mathbf{L} defined as follows:

$$(\mathbf{L})_{ij} = \ell_{(i,j)} \quad i, j = 1, \dots, n \quad (6)$$

This matrix is shown in Fig. 2.a for the two time-series of concern. Starting from the upper left and ending in the lower right corners, the sequence of cells marked with heavy lines corresponds to the warping path shown

in Fig. 2.c. It is easily checked that this path obeys the continuity and monotonicity constraints. The latter allow passing from cell (i, j) to cell (i', j') through forward-forward ($i' = i + 1, j' = j + 1$), delay-forward ($i' = i, j' = j + 1$), and forward-delay ($i' = i + 1, j' = j$) moves. The corresponding distance D is the sum of the \mathbf{L} matrix entries visited by the warping path.

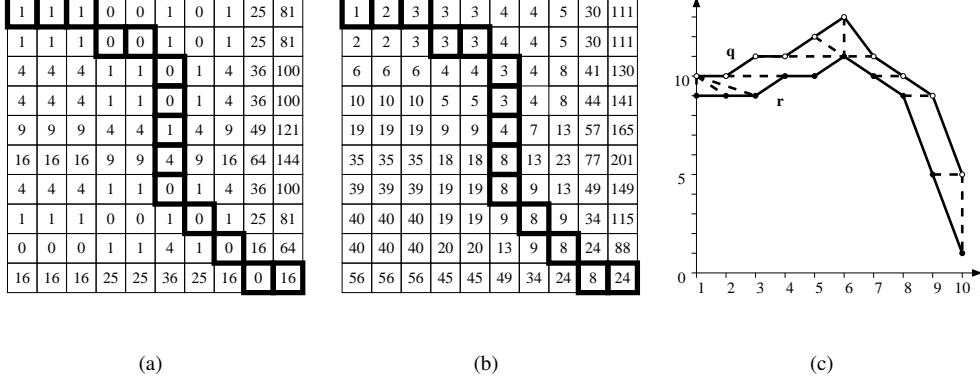


Figure 2: Classical DTW distance: local distance matrix \mathbf{L} (a), accumulated distance matrix Δ (b), query and reference plots with warped pairs (c)

To determine “how close” the time-series \mathbf{q} and \mathbf{r} are from each other, it is appropriate to consider the warping path which minimizes $D_d(\mathbf{q}, \mathbf{r}, \mathbf{w})$, i.e. to determine the \mathbf{w}^* solution of:

$$\min_{\mathbf{w}} D(\mathbf{q}, \mathbf{r}, \mathbf{w}) = \min_{\mathbf{w}} \sqrt{\sum_{k=1}^N \ell_{w_k}} \quad (7)$$

and use the corresponding distance $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*)$ as the sought measure. The latter will be referred to as the DTW distance.

Note that if one imposes $N = n$, the unique warping path is $\mathbf{w} = [(1, 1) (2, 2) \dots (n, n)]$, which lies along the diagonal of \mathbf{L} , and $D(\mathbf{q}, \mathbf{r}, \mathbf{w})$ coincides with the Euclidean distance $E(\mathbf{q}, \mathbf{r})$, which is also the square root of the trace of the \mathbf{L} matrix.

Coming back to the example, the path emphasized in Fig. 2.a turns out to be the optimal warping path \mathbf{w}^* . One could check that no other association between the points of \mathbf{q} and \mathbf{r} (satisfying constraints 1 to 3) than the one shown with dotted lines in Fig. 2.c yields a smaller value of $D(\mathbf{q}, \mathbf{r}, \mathbf{w})$. The corresponding DTW distance is:

$$D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*) = \sqrt{(3 \times 1) + (4 \times 0) + 1 + 4 + (4 \times 0) + 16} = \sqrt{24} = 4.899$$

while the Euclidean distance is $E(\mathbf{q}, \mathbf{r}) = \sqrt{49} = 7$.

With reference to Fig. 2.c, it can be said that the optimal warping path defines the deformation of the time axes of \mathbf{q} and \mathbf{r} which brings the two time-series as close as possible to each other (in the Euclidean-norm sense) [4].

2.2 Determination of the optimum warping path

Finding the optimal warping path is a combinatorial optimization problem. In spite of the large search space, this problem can be solved with remarkable efficiency through Dynamic Programming (DP) [2, 3].

DP exploits the fact that each portion of the optimal warping path is itself an optimal warping path. More precisely the DP-based algorithm relies on the recursive formula:

$$D(\mathbf{q}, \mathbf{r}, \mathbf{w}_{1 \rightarrow k}^*) = \ell_{w_k} + D(\mathbf{q}, \mathbf{r}, \mathbf{w}_{1 \rightarrow k-1}^*) \quad (8)$$

$$\text{with } D(\mathbf{q}, \mathbf{r}, \mathbf{w}_{1 \rightarrow 1}^*) = \ell_{w_1} = \ell_{(1,1)} \quad (9)$$

where $\mathbf{w}_{1 \rightarrow k}^*$ denotes the first k components of \mathbf{w}^* .

In practice, it is convenient to consider the accumulated distance matrix Δ defined recursively as:

$$[\Delta]_{1j} = \sum_{k=1}^j [\mathbf{L}]_{1k} \quad j = 1, \dots, n \quad (10)$$

$$[\Delta]_{i1} = \sum_{k=1}^i [\mathbf{L}]_{k1} \quad i = 2, \dots, n \quad (11)$$

$$[\Delta]_{ij} = [\mathbf{L}]_{ij} + \min\{[\Delta]_{i-1,j-1}, [\Delta]_{i,j-1}, [\Delta]_{i-1,j}\} \quad i, j = 2, \dots, n \quad (12)$$

This matrix is built starting from the upper left corner. From (5, 7, 8, 9) it is easily shown that:

$$D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*) = \sqrt{[\Delta]_{nn}} \quad (13)$$

Coming back again to the example, Fig. 2.b shows the Δ matrix computed by (10-12) from the \mathbf{L} matrix shown in Fig. 2.a. Furthermore, the optimal warping path is identified by backtracking from the (n, n) entry of Δ , taking as predecessor of the (i, j) cell, the one among $\{(i-1, j-1), (i, j-1), (i-1, j)\}$ with the lowest value, and repeating the procedure until the $(1, 1)$ cell is reached. It is easily checked that this simple procedure yields the path emphasized in Fig. 2.b.

Even if the computational cost of the DTW algorithm is not an issue for the off-line accuracy studies considered here, it may be of interest to reduce the computational burden of the DTW algorithm by using the technique proposed in [3]. The idea is to restrain the warping path from regions of the local distance matrix \mathbf{L} which are

far from its diagonal. Thus, the warping path is searched in what has been eponymously called the Sakoe-Chiba band. In speech recognition some heuristics are used to choose the width of the band, which it would probably not make sense to transpose to power systems.

Instead, that width can be related to the largest acceptable time difference τ_{max} between any two associated points (i.e. the largest acceptable time warping). When the query and reference curves span the same time interval T and have n uniformly distributed sampling points, separated by the time step $h = T/(n - 1)$, this leads to computing only a band of m points on both sides of the diagonal of \mathbf{L} , where $m = \tau_{max}/h$.

Figure 3 shows the case where m has been set to 5 in the example of Fig. 2. Since the unconstrained optimal warping path was not visiting the black cells of \mathbf{L} , this path is identical to that in Fig. 2. So is the DTW distance $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*)$.

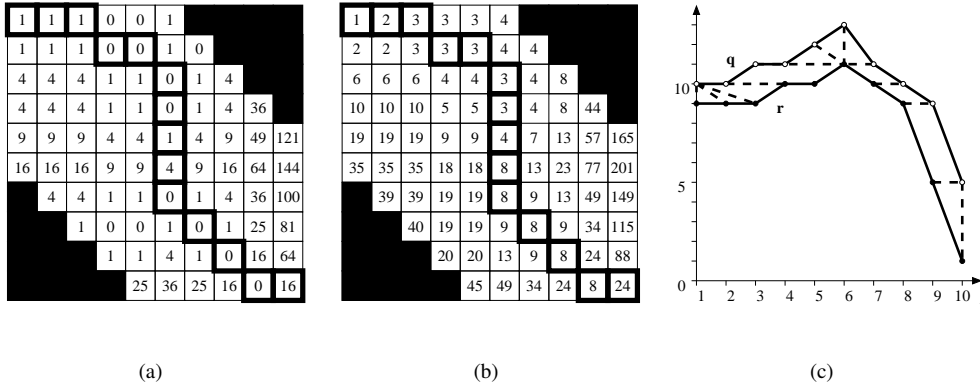


Figure 3: DTW with Sakoe-Chiba band: local distance matrix \mathbf{L} (a), accumulated distance matrix Δ (b), query and reference plots with warped pairs (c)

In this work, the Sakoe-Chiba band was not used to bound the maximum time difference between any two associated points, but only to avoid useless computations in regions of \mathbf{L} that are not likely to be visited by any optimal path. For instance, for long curves spanning a time interval T of a few minutes, τ_{max} was set to 30 seconds. Clearly, this choice is problem dependent and some tuning is required if computational effort is an issue.

2.3 Practical uses of DTW distance and optimal warping path

Although it makes sense to minimize $D(\mathbf{q}, \mathbf{r}, \mathbf{w})$ to identify the optimal warping path between two time-series, the value of $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*)$ by itself may not be a meaningful measure of their proximity. For instance, if a larger number of points is used for the comparison, the value of $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*)$ will increase accordingly. To deal with this situation, it sounds more appropriate to consider the normalized DTW distance:

$$\delta = \sqrt{\frac{\sum_{k=1}^N \ell_{w_k}}{N}} = \frac{D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*)}{\sqrt{N}} \quad (14)$$

If the two curves are smooth enough compared to the sampling period, a higher sampling will lead to a higher value of N but one can expect δ to be few affected.

Further practical information can be obtained from the optimal warping path. With the pairs of indices in \mathbf{w}^* denoted as in (4), we define the vector of curve offsets as:

$$\mathbf{y} = [q_{i_1} - r_{j_1}, q_{i_2} - r_{j_2}, \dots, q_{i_N} - r_{j_N}] \quad (15)$$

and the vector of time shifts as:

$$\mathbf{x} = [t_{i_1} - t_{j_1}, t_{i_2} - t_{j_2}, \dots, t_{i_N} - t_{j_N}] \quad (16)$$

The differences between the two curves can be measured by:

- the average curve offset:

$$\mu = \frac{\sum_{k=1}^N [\mathbf{y}]_k}{N} = \frac{1}{N} \sum_{k=1}^N (q_{i_k} - r_{j_k}) \quad (17)$$

- the standard deviation of the curve offsets:

$$\sigma = \sqrt{\frac{\sum_{k=1}^N ([\mathbf{y}]_k - \mu)^2}{N}} \quad (18)$$

- the average time shift:

$$\tau = \frac{\sum_{k=1}^N [\mathbf{x}]_k}{N} = \frac{1}{N} \sum_{k=1}^N (t_{i_k} - t_{j_k}) \quad (19)$$

For instance, when comparing two time-series, a small value of μ indicates that, with a proper deformation of the time axes, the curves coincide on the average. τ characterizes the average delay corresponding to that deformation. When comparing several queries to a reference, a larger value of σ denotes more ‘‘volatility’’ with respect to that reference. In case of oscillatory responses, for instance, a small σ indicates that both curves have similar damping.

3 Improvements of the DTW distance

3.1 Extending the method to open-end DTW

The DTW method was initially developed for speech recognition applications where the signals to be compared start and end at zero values. In power system applications where simulations (or simulations and measurements) have to be compared, it is reasonable to assume that the two curves start from steady-state values. However, they may eventually evolve into system instability, in which case they will end in the middle of transients caused by either a monotonic collapse or by growing oscillations.

In such situations, it may not be meaningful to impose the optimal warping path to involve all the n points of each curve. Indeed, under the effect of the above mentioned transients, $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*)$, μ or σ may be significantly and falsely influenced by the last points of the curves. This is even more true in a collapse simulation where the model accuracy is questionable in those degraded operating conditions and, hence, less importance should be given to the final values.

This motivates a modification of the method referred to as open-end DTW, which gives better performances than the standard DTW algorithm [4], allowing only a partial alignment of the time-series thanks to a softening of the Boundary condition defined in Section 2.1.

Thus, instead of requesting the optimal warping path to end up with $w_N = (i_N, j_N) = (n, n)$, it is allowed to end up in $w_N = (i_N, j_N)$ with:

$$\begin{aligned}n' - m' &\leq i_N \leq n \\n' - m' &\leq j_N \leq n \\(j_N - n)(i_N - n) &= 0\end{aligned}$$

where the equality forces the last point of at least one of the two curves to be involved in the DTW distance.

Figure 4 shows an example of open-end DTW. In this case m' has been set 5. As a result, the optimal warping path stops in the (10, 9) entry of \mathbf{L} instead of (10, 10) as in Figs. 2 and 3. It is easily seen from Fig. 4.c that from $t = 6$ to $t = 10$, the query curve is delayed with respect to the reference, and therefore does not go as low as the latter, although the “plunging” evolution is reproduced. The optimal warping path does not involve the last point of the reference curve. This leads to a the DTW distance $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*) = \sqrt{8} = 2.828$. This value is significantly

smaller than the DTW distance of 4.899 obtained under strict boundary condition. This smaller distance reflects the better matching between both curves when ignoring the last point of the reference time-series.

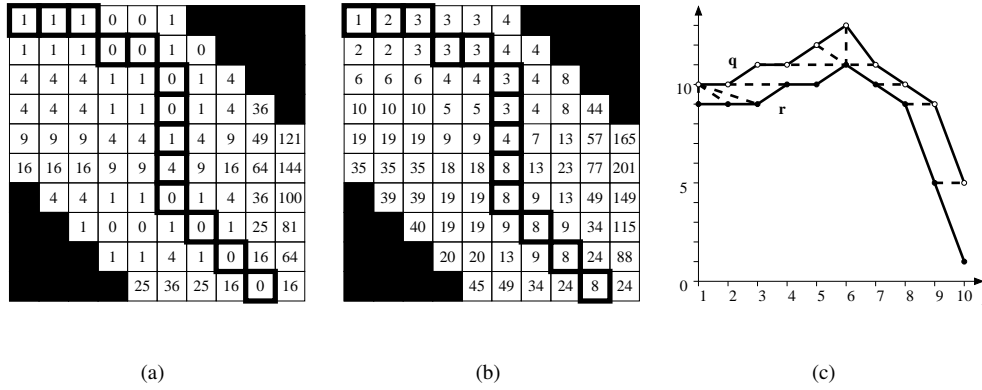


Figure 4: Open-end DTW: local distance matrix L (a), accumulated distance matrix Δ (b), query and reference plots with warped pairs (c)

Similarly to m , and re-using the notation of Section 2.2, m' can be taken as $m' = \tau'_{max}/h$ where τ'_{max} is the largest acceptable time difference on the last pair of associated points.

3.2 Penalizing large time warps

It has been noticed that on some occasions, the DTW algorithm “over-warps” the time axes during its optimization process. This happens in the form of a single point of one time-series being mapped to a large number of points of the other time-series, and was referred to as a singularity in [5]. In Figs. 2, 3 and 4, for instance, the 6th point of the reference curve is a significant singularity in so far as it is mapped to 5 points of the query curve.

Several methods have been proposed to alleviate this problem [5]. In this work, a modification of the slope-weighting technique described in [3] was found sufficient to remove or decrease the importance of singularities.

A singularity corresponds to several steps taken successively in the delay-forward (respectively the forward-delay) direction. Let us denote by k_{df} the number of steps taken consecutively in the delay-forward direction and by k_{fd} its counterpart in the forward-delay direction. k_{df} is reset to zero as soon as a forward-forward or a forward-delay step is taken, and similarly for k_{fd} . Equations (12-11) used to compute the accumulated distance

matrix are modified as follows:

$$[\Delta]_{1j} = \sum_{k=1}^j p^{k-1} [\mathbf{L}]_{1k} \quad j = 1, \dots, n \quad (20)$$

$$[\Delta]_{i1} = \sum_{k=1}^i p^{k-1} [\mathbf{L}]_{k1} \quad i = 2, \dots, n \quad (21)$$

$$[\Delta]_{ij} = \min\{[\Delta]_{i-1,j-1} + [\mathbf{L}]_{ij}, [\Delta]_{i,j-1} + p^{k_{df}} [\mathbf{L}]_{ij}, [\Delta]_{i-1,j} + p^{k_{fd}} [\mathbf{L}]_{ij}\} \quad i, j = 2, \dots, n \quad (22)$$

where $p \geq 1$. Thus, the $p^{k_{df}}$ and $p^{k_{fd}}$ factors penalize successive mappings that involve a single point of the same curve. The choice of p may require some tuning as it is problem and sampling dependent. As a rule of thumb, in case of a uniform sampling with time step h , and assuming that a singularity with cumulated delay Δ_d is to be penalized with a factor $P \geq 1$, the recommended value is :

$$p = P^{h/\Delta_d}$$

Note that the larger p , the closer the optimal warping path to the diagonal of Δ and hence, the closer the DTW distance to the Euclidean one.

Figure 5 shows an example of application of the slope-weighted DTW. In this example, the ‘‘large’’ value $p = 2$ was chosen to highlight the slope-weighting effect: as expected, it yields a nearly-diagonal warping path. The mapping of points is shown in Fig. 5. Under the effect of penalties, the singularity of the 6th point of the reference curve has decreased, since it is now mapped to only two points of the query curve. The corresponding DTW distance is $\sqrt{\sum_{k=1}^N \ell_{w_k}} = 3.873$.

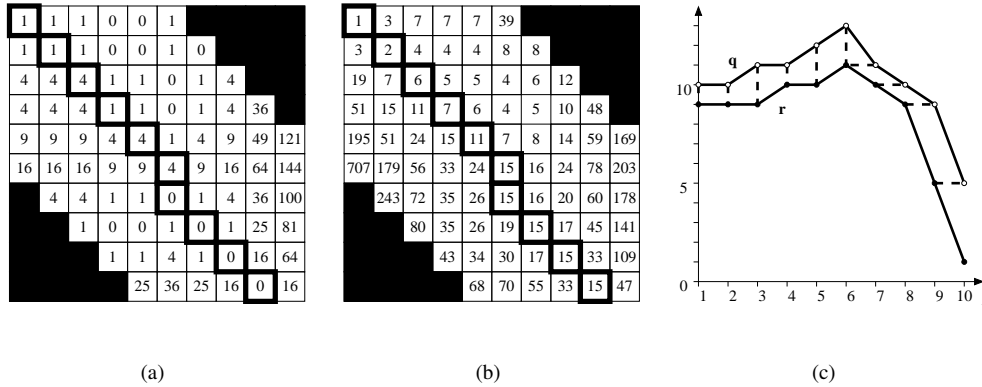


Figure 5: Slope-weighted DTW: local distance matrix \mathbf{L} (a), accumulated distance matrix Δ (b), query and reference plots with warped pairs (c)

4 Discussion

4.1 Metric aspects

A desirable property for a measure of distance is that it is a “metric”, which requires to satisfy the following four conditions (the dependency on \mathbf{w}^* is dropped for simplicity of writing):

- *Non-negativity*: $D(\mathbf{q}, \mathbf{r}) \geq 0$.
- *Symmetry*: $D(\mathbf{q}, \mathbf{r}) = D(\mathbf{r}, \mathbf{q})$.
- *Identity of indiscernibles*: $D(\mathbf{q}, \mathbf{r}) = 0 \iff \mathbf{q} = \mathbf{r}$.
- *Triangle inequality*: $D(\mathbf{q}, \mathbf{q}') + D(\mathbf{q}', \mathbf{r}) \geq D(\mathbf{q}, \mathbf{r})$.

It is easily checked that the first two properties hold true for the DTW distance. On the other hand, it is easily verified with counterexamples that it does not satisfy the last two properties.

In the speech recognition problem, the task is to compare the query (a spoken word) with many references (all word prototypes in a dictionary) to identify the closest to the query (the most likely spoken word). In that problem and, in general, in search algorithms, a measure with metric properties would allow finding the sought reference without computing the DTW distances between the query and all references, which could be computationally demanding, especially with a large number of references.

For the power system applications reported in this paper, however, the absence of a metric structure is not problematic at all, since our purpose is somewhat opposite: determine which of the queries is the closest approximation of a single reference.

While no study has been carried on about loose satisfaction of metric properties in the power system case, empirical evidence of quasi-metric properties was given in a speech recognition problem [10].

4.2 Sampling effects

A possible weakness of the DTW distance is its dependence on the curve sampling. Different samplings may lead to quite different results. For instance, consider the following continuous-time evolutions:

$$q(t) = \text{floor}(t) \quad (\text{i.e. the largest integer } \leq t)$$

$$r(t) = t$$

sampled to get the time-series \mathbf{q} and \mathbf{r} . Clearly, with an integer sampling period, starting at $t = 0$, $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*) = 0$ while with any other sampling $D(\mathbf{q}, \mathbf{r}, \mathbf{w}^*) \neq 0$.

Algorithms for continuous dynamic time warping have been proposed [6] to deal with this problem, at the cost of increased complexity and computational effort. This paper sticks with the classical, discrete formulation to preserve the intuitiveness and simplicity of the DTW algorithm. The impact of sampling can be reduced by using a small enough sampling period. The resulting computational burden is compensated by storing and processing only the Sakoe-Chiba band in \mathbf{L} and Δ .

4.3 Multidimensional DTW

So far it has been considered that the curve comparison involves a set of queries and a single reference. However, since power systems are multi-dimensional systems, it might be of interest to compute a global distance measure between several sets of queries and their corresponding references. To this purpose the concept of multi-dimensional DTW is introduced.

Let \mathcal{S} be a set of variables of interest. Note that \mathcal{S} should not be too large mainly because involving too many curves exposes to masking effects. For instance, when analyzing the impact of a disturbance on a large power system, it is likely that components far away from its location will respond comparatively less. Involving them in the comparison may “dilute” significant discrepancies on the most affected variables, and potentially give a false impression of accuracy.

For the sake of simplicity, we still suppose that the query \mathbf{q}^k and reference \mathbf{r}^k time-series of the same k -th variable in \mathcal{S} span the same time interval and have the same number n of uniformly sampled points.

The squared distance between two points of respectively \mathbf{q}^k and \mathbf{r}^k is defined, analogously to (2), as:

$$\ell_{(k,i,j)} = (q_i^k - r_j^k)^2 \quad i, j = 1, \dots, n \quad (23)$$

The local distance matrix \mathbf{L} is redefined as follows:

$$(\mathbf{L})_{ij} = \sum_{k \in \mathcal{S}} w_k \ell_{(k,i,j)} \quad i, j = 1, \dots, n \quad (24)$$

where w_k is the weight associated to the k -th variable. This weight may account for a conversion between different units (e.g. kV and MW) and/or different importance given to different variables (higher w_k corresponding to

higher importance). From there on, the remaining of the procedure detailed in Sections 2.2 and 3, is identical. Therefore, the additional computational effort with respect to the single-dimensional case is very limited, since it has to do with the construction of the \mathbf{L} matrix only.

5 Simulation results

This section illustrates the DTW method (and its extension) in two typical applications.

5.1 Solver accuracy comparison

In the first application, the DTW method is used to measure the accuracy of fast, simplified simulation methods.

The case has been obtained with the Nordic32 test system documented in [13]. In the scenario of concern, following a line outage, the system undergoes long-term voltage instability under the effect of overexcitation limiters reducing the field currents of generators and load tap changers attempting to restore the distribution voltage (and hence load powers).

The benchmark evolution has been obtained with the 2nd-order Backward Differentiation Formula (BDF) integration formula, using a step size of 0.05 s. Two simplified simulations have been considered:

- fast solver: following the ideas in [11], the backward Euler method has been used to simulate the same detailed model as in the benchmark but with a large step size of 0.5 s and an appropriate handling of discrete events as explained in [12]. This approach is denoted by BE in the sequel;
- model simplification: the quasi steady-state approximation of the long-term dynamics has been considered [9], which consists of decomposing the model into fast and slow dynamics, replacing the former by their equilibrium conditions and integrating the latter with a large step size of 1 s. This yields the fastest but also the more approximate time simulation. This approach is denoted by QSS in the sequel.

Figure 6.a shows the time evolution of a representative bus voltage obtained by the benchmark (solid lines) and the two simplified (dashed and dotted, respectively) simulation methods. A zoom on the first 10 s is provided in Fig. 6.b. As can be seen, BE overlooks some of the oscillations but matches the benchmark once the large transients have died out. QSS reproduces the same pattern but with a delay that increases towards the end.

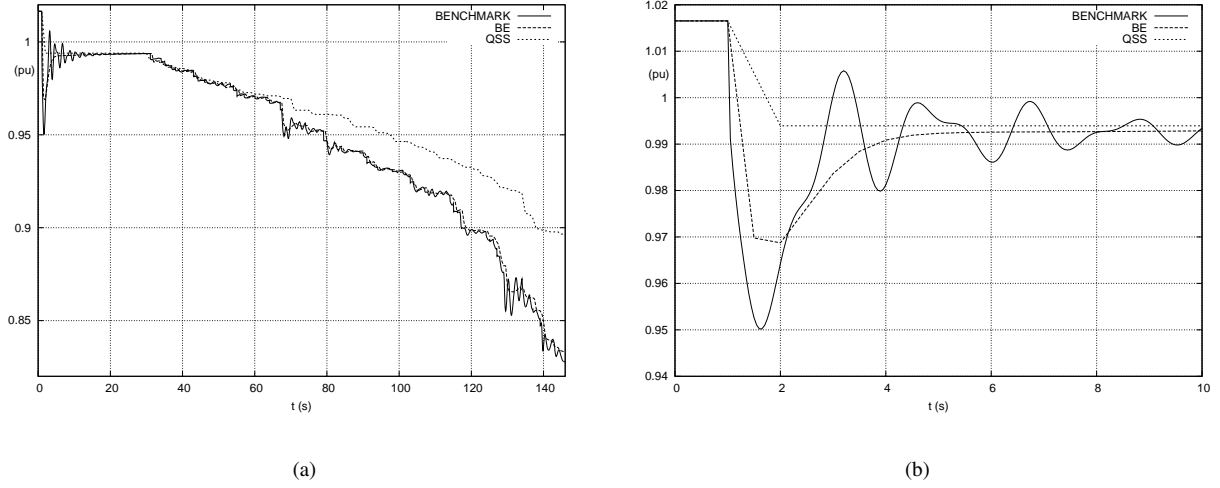


Figure 6: Voltage instability scenario: long-term evolution of voltage (a) and zoom over the first 10 s (b)

To compute the DTW distance, all curves have been sampled every 0.05 s, which is the smallest integration step size (used to obtain the reference curve). Linear interpolation has been used to sample the BE and QSS curves. For easier comparisons with DTW, we consider the normalized Euclidean distance:

$$\epsilon = \frac{E(\mathbf{q}, \mathbf{r})}{\sqrt{n}} \quad (25)$$

This distance is found to be $\epsilon = 3.499 \times 10^{-3}$ pu for BE and $\epsilon = 24.45 \times 10^{-3}$ pu for QSS, which would suggest that QSS is nearly one order of magnitude less accurate than BE.

The DTW distances show that in fact the QSS simulation is not so inaccurate, but just suffers from a growing delay, as can be concluded from a visual inspection of the curves. The DTW measures have been computed with $\tau_{max} = \tau'_{max} = 30$ s (which corresponds to $m = m' = 600$) and $p = 1.5$. The results are summarized in Table 1.

Table 1: solver accuracy comparison; collapse scenario (all values in pu of voltage)

solver	δ	μ	σ	τ
BE	1.876×10^{-3}	0.132×10^{-3}	1.872×10^{-3}	0.769
QSS	3.537×10^{-3}	0.495×10^{-3}	3.503×10^{-3}	11.19

The values of δ show that QSS is only twice as inaccurate as BE. μ indicates that QSS suffers from a positive offset larger than BE. The BE curve is characterized by a smaller value of σ because it reproduces part of the

initial voltage dip while the QSS curve goes directly to the underlying equilibrium, as can be seen in Fig. 6.b. τ confirms that QSS suffers from a large average delay.

The simplified solvers considered here are not aimed at reproducing the electromechanical oscillations [11]. However, a slightly more accurate simulation can be obtained - with little additional computational effort - by resorting to the second-order BDF integration scheme instead of Backward Euler (which is first-order BDF), while still keeping the large step size of 0.5 s. This approach is denoted by BDF2 in the sequel.

The respective accuracies of BE and BDF2 with respect to the benchmark have been determined for the curves in Fig. 7, showing the short-term evolution of the active power produced by a generator, in response to another line outage in the Nordic32 system. As can be seen, BDF2 partly reproduces the oscillations while BE just sketches the first swing.

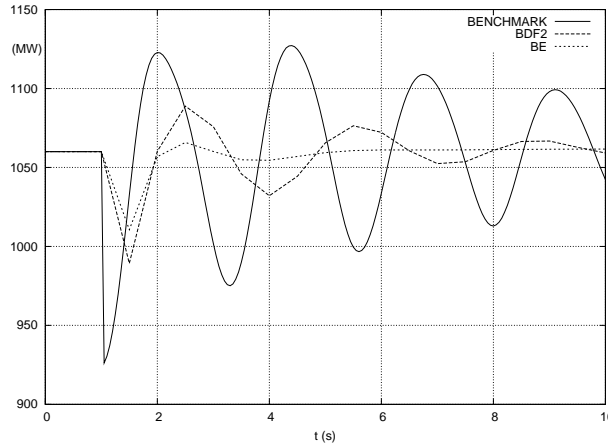


Figure 7: Oscillatory responses to be compared

The normalized Euclidean distances defined by (25) are $\epsilon = 47.23$ MW for BDF2 and $\epsilon = 43.65$ MW for BE which would suggest that both solvers are quite inaccurate, and paradoxically, BDF2 is the least accurate.

The DTW measures have been computed with $\tau_{max} = \tau'_{max} = 5$ s (which corresponds to $m = m' = 100$) and $p = 1.5$. The results, in MW, are given in Table 2. Unlike the Euclidean distances, the δ values indicate that BDF2 is more accurate, which confirms the observation that, in spite of its high damping, BDF2 still reproduces an oscillation, which is overlooked in the BE response. The μ values indicate that both solvers suffer from a negative offset, while the σ values show that BDF2 has damping closer to the reference than BE. The τ values indicate that BE suffers from an average delay twice as large as BDF2.

Table 2: Solver accuracy comparison; oscillatory scenario (all values in MW)

solver	δ	μ	σ	τ
BDF2	27.94	-1.085	27.99	1.077
BE	34.09	-0.547	34.17	2.043

5.2 Model accuracy comparison

We now illustrate how the DTW method can be used to validate the outputs of models with respect to time responses measured on the real system.

The case relates to a three-phase fault that occurred in 2010 on the 315-kV network of Hydro-Québec, when energizing a transformer. The fault was cleared in 6 cycles, while some load spontaneously tripped after 3 cycles. The recorded voltage evolution at a 735-kV bus is shown with solid line in Fig. 8. The sampling period is 1 cycle at 60 Hz ($\simeq 0.0167$ s).

Two models were considered when reproducing the event with an industrial software:

- loads wholly represented by a static, exponential model. This model is referred to as CS in the sequel
- part of the load represented by an equivalent induction motor. This model is referred to as CM in the sequel.

The simulated response obtained with each model, using a time step size of 0.25 cycle ($\simeq 0.0042$ s) is shown in Fig. 8.a with dashed and dotted lines, respectively. Figure 8.b provides a zoom on the first 3 s.

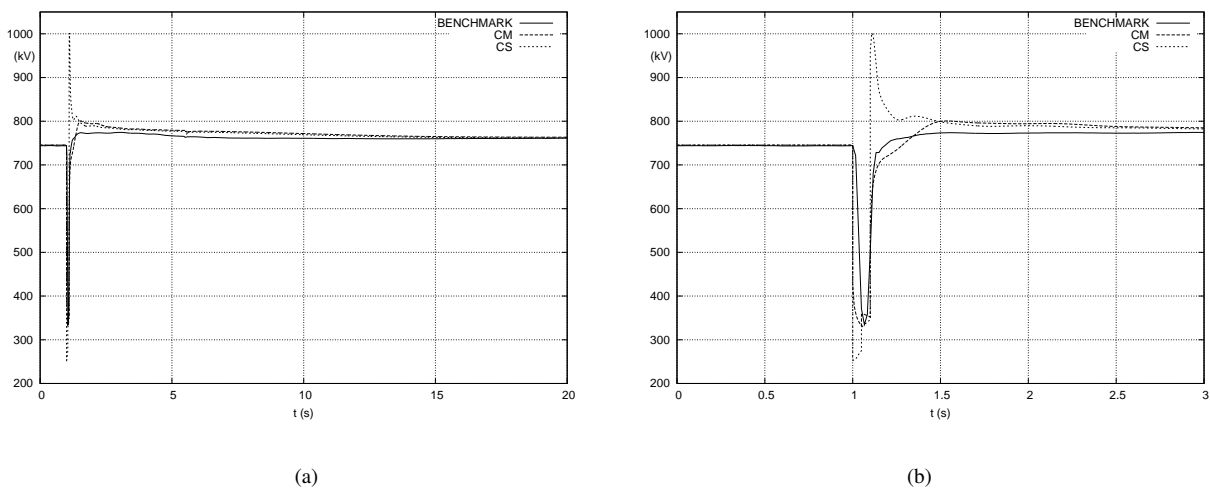


Figure 8: Analysis of a real incident: evolution of voltage at a 735-kV bus (a) and zoom over the first 3 s (b)

Clearly, the CS model yields a large voltage spike after fault clearing while the CM model gives results closer to the measurements. Both simulations join the benchmark once the large transients have died out.

To further measure the accuracy of the two models, all curves have been sampled with a period of 0.25 cycle and with linear interpolation. The normalized Euclidean distances in kV are $\epsilon = 17.59$ kV for CM and $\epsilon = 26.89$ kV for CS.

The DTW distances have been computed with $\tau_{max} = \tau'_{max} = 2.5$ s (which corresponds to $m = m' = 600$) and $p = 1.035$. The results are given in Table 3.

Table 3: Model accuracy comparison (all values in kV)

Model	δ	μ	σ	τ
CM	9.134	6.806	6.093	1.877
CS	18.17	5.931	17.18	1.976

The respective values of δ show that CM is twice as accurate as CS. μ indicates that both models suffer from a positive offset. σ shows that CM has a damping closer to the reference than CS. τ indicates that both models have a delay around 2 s.

6 Conclusion

A measure of proximity of time-series stemming from power system simulations and/or measurements has been proposed in this paper. This measure is based on Dynamic Time Warping, an algorithm originally developed for speech recognition and now used in several fields. This algorithm warps the time axis and maps the points of both curves to minimize the sum of their squared ordinate differences. This is equivalent to determining the deformation of the time axes which brings the two time-series as close as possible in the Euclidean-norm sense. An efficient dynamic programming algorithm is used for the minimization. Extensions and adaptations of the classical method, to deal with typical power system responses have been presented, together with statistics allowing to characterize the proximity of two time evolutions with as few as 3 parameters: μ , τ and σ . A key point

of any proposed metric is how well it correlates with expert opinion. A further validation of the DTW approach could be obtained by comparing the DTW-based ranking of a set of queries with the ranking made by a power system expert. Depending on the application, the most significant ranking would be based on μ , τ or σ . A multi-dimensional extension allowing system-wide measures of similarity has been also discussed. The effectiveness and usefulness of the method has been demonstrated on several power system applications. The relative accuracy of a set of simplified simulations with respect to a reference was computed on a small system. In the context of model validation, two simulated responses were compared to a recorded measurement of a real incident.

Acknowledgement

This work was performed in the context of the PEGASE project [14] funded by European Community's 7th Framework Programme (grant agreement No. 211407).

References

- [1] Schwer, L.E.: 'Validation metrics for response histories: perspectives and case studies', *Engineering with Computers*, Oct. 2007, Vol. 23, Issue 4, pp. 295-309
- [2] Vintsyuk, T.K.: 'Speech discrimination by dynamic programming', *Kibernetika (in Russian)*, 1968, Vol. 4, No. 1, pp. 81-88
- [3] Sakoe, H., and Chiba, S.: 'Dynamic programming algorithm optimization for spoken word recognition', *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Feb. 1978, Vol. 26, No. 1, pp. 43- 49
- [4] Tormene, P., Giorgino, T., Quaglini, S., and Stefanelli, M.: 'Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation', *Artificial Intelligence in Medicine*, Jan. 2009, Vol. 45, Issue 1, pp. 11-34
- [5] Keogh, E.J., and Pazzani, M.J.: 'Derivative dynamic time warping', *Proc. 1st SIAM International Conference on Data Mining*, Chicago, USA, Apr. 2001, pp. 1-11

- [6] Efrat, A., Fan, Q., and Venkatasubramanian, S.: 'Curve matching, time warping, and light fields: new algorithms for computing similarity between curves', *Journal of Mathematical Imaging and Vision*, Apr. 2007, Vol. 27, No. 3, pp. 203-216
- [7] Youssef, A.M., Abdel-Galil, T.K., El-Saadany, E.F., and Salama, M.M.A.: 'Disturbance classification utilizing dynamic time warping classifier' *IEEE Trans. Power Delivery*, Jan. 2004, Vol. 19, No. 1, pp. 272-278
- [8] Kong, Y., and Liu, S.: 'Power System Transient Stability Assessment Based on Geometric features of Disturbed Trajectory and DTW', *Proc. 6th International Conference on Natural Computation (ICNC 2010)*, Yantai, China, Aug. 2010, pp. 485-489
- [9] Van Cutsem, T., and Vournas, C.: 'Voltage Stability of Electric Power Systems' (Kluwer Academic Publishers - now Springer, 1998)
- [10] Vidal Ruiz, E., Casacuberta Nolla, F., and Rulot Segovia, H.: 'Is the DTW distance really a metric? An algorithm reducing the number of DTW comparisons in isolated word recognition', *Speech Communication*, Dec. 1985, Vol. 4, Issue 4, pp. 333-344
- [11] Fabozzi, D., and Van Cutsem, T.: 'Simplified time-domain simulation of detailed long-term dynamic models', *Proc. IEEE PES General Meeting, Calgary, Canada, July 2009*, pp. 1-8. Available at <http://ieeexplore.ieee.org> (DOI 10.1109/PES.2009.5275463) and <http://hdl.handle.net/2268/9524>
- [12] Fabozzi, D., Chieh, A.S., Panciatici, P., and Van Cutsem, T.: 'On simplified handling of state events in time-domain simulation', accepted for presentation at the 17th Power System Computation Conference (PSCC), Stockholm, Sweden, Aug. 2011, pp. 1-9. Available at <http://hdl.handle.net/2268/91650>
- [13] Glavic, M., and Van Cutsem, T.: 'Wide-Area Detection of Voltage Instability From Synchronized Phasor Measurements. Part II: Simulation Results', *IEEE Trans. Power Syst.*, Aug. 2009, Vol. 24, No. 3, pp. 1417-1425
- [14] <http://www.fp7-pegase.eu/>, accessed May 2011