# Let's Talk About DMA Size

*The pros and cons of dividing a water distribution network into **District Metered Areas** (DMAs) are well understood. But how can you tell what size DMAs will best enable you to locate leaks?*

*Let's look at 15 water utilities from around the world and more than 1000 DMAs of different shapes and sizes and find out!*

*We will learn: (1) The kind of tradeoffs to consider when choosing DMA sizes; (2) the most effective DMA size range for finding leaks is between 500 and 2000 connections; (3) that there is an empirical recipe to assess the minimal detectable leaks.*

## No Such Thing as One-Size-Fits-All

We want to know about every ongoing leakage in our water distribution systems. This knowledge is critical for planning the perfect chain of repair operations. Surprisingly, we *can* obtain perfect knowledge in a very methodical way – by making sure that our water distribution network architecture is covered by a **swarm of *small* DMAs**. This approach makes it possible to calculate the water supply in each small region, providing a very cheap way of zeroing in on trends, anomalies, creeping nightlines, bursts, etc. This is achieved without installing meters on every pipe, thanks to the math of mass conservation rules. These rules state that the total water supply in a region equals the incoming volume minus the outgoing volume. Analytical solutions, such as *TaKaDu*, help a lot in organizing the supply information into a prioritized flow of actions to be taken.

On the other hand, there is also merit to the extreme opposite network architecture. It is possible to get away with designing **fat bulky DMAs** or even not having DMAs at all. Having no small DMAs increases redundancy of pipe-paths (which is good), limits dead ends with reduced water quality, and is more energy efficient. Such architecture is also a lot cheaper to create (but is not necessarily cheaper to maintain).

Therein lies the tension: you can get perfect information, but you pay for it in other ways.

DMAs should not be too big or small. **If too small,** your supply calculations will be very unstable (more on this, below). **If too big,** you will not be able to detect small anomalies or leaks, only very large ones. So (for both DMA-based detection and in general), we want to know what is the ***smallest detectable leak*** when using data analytics. If you can detect leaks of 0.1 l/s or even smaller[1], then you are probably OK. When utilizing DMAs for leak detection, the ***smallest detectable leak*** is determined mostly by statistics. We will discuss this point below.


The good news is that there's a lot of wiggle room in the range between too-small and the too-big. In the next section, you will be able to locate your organization on this spectrum by using three straightforward metrics for DMA size.

## How Big is Big? How Small is Small?

There are three simple metrics to quantify DMA sizes:
(1) The **number of connections**

[1] This is the theoretical minimum. In real-life data, a minimum leak size that is detectable is 0.5 l/s, depending on data noise.

(2) The distribution pipes' **total length**
(3) The average **daily supply**

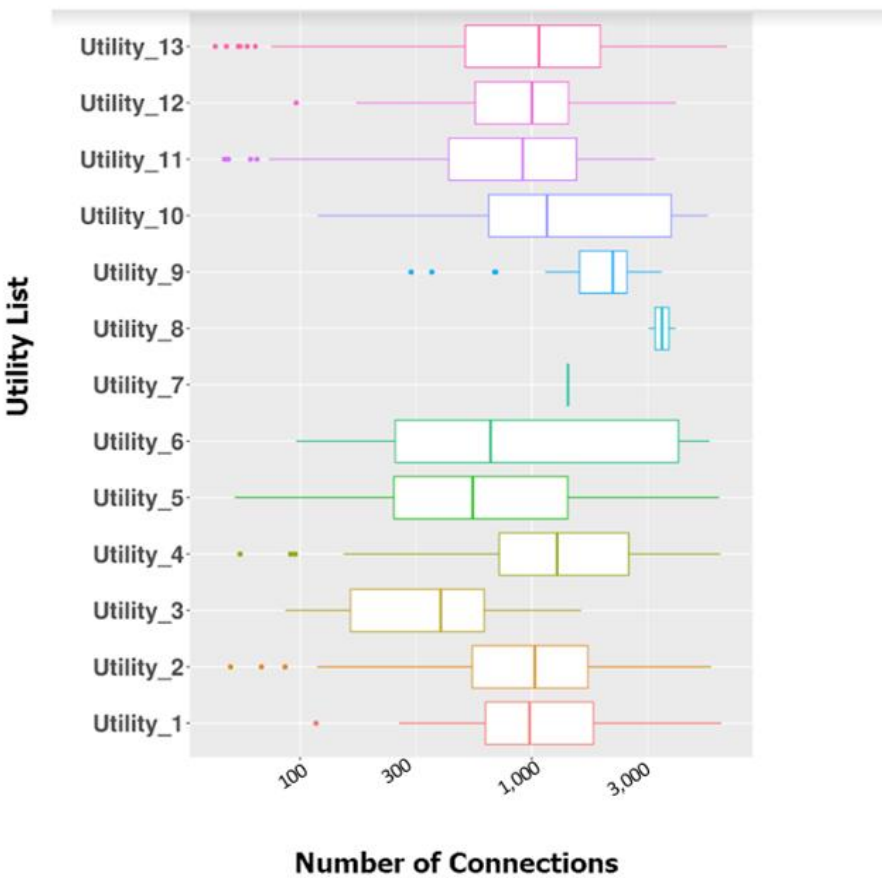
Often, the daily supply is not available during the planning stage, without using expensive simulations. Below, we will give you a recipe for obtaining any one of these metrics if you have the other two. Before doing so, you may want to know where your water utility stands with respect to the metrics that you do have at hand.

Through our work with dozens of water utilities around the world, we have gathered extensive statistics covering large and small water utilities in numerous countries and continents, all long-time users of TaKaDu.

In Figures 1-3, you can see the various DMA sizes in each utility, represented by box plots. Each box shows the middle 50% of DMAs for each water utility. The rest of the DMAs are shown in the extended whiskers.

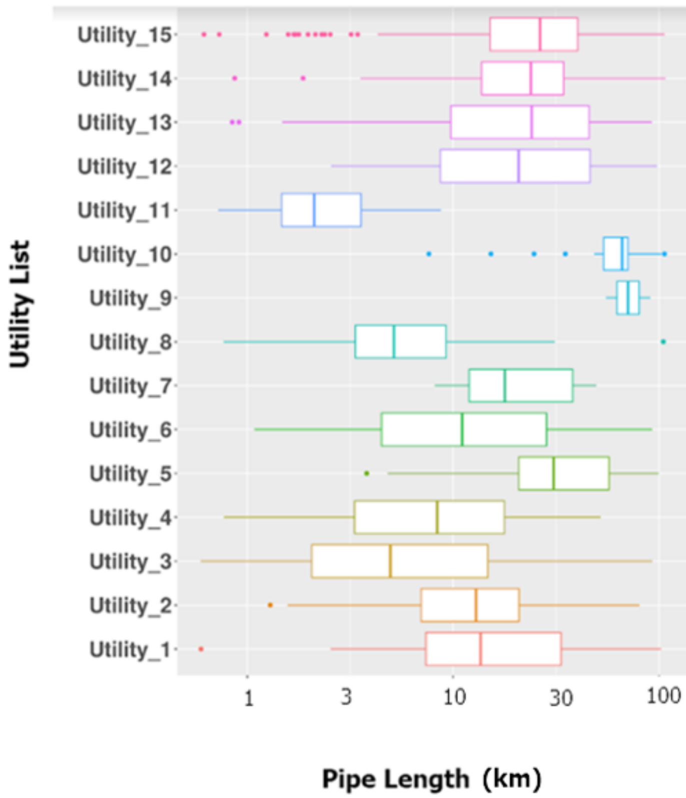
In Figure 1, we see both the *swarm* approach (e.g., in Utility 3), as well as the *fat bulky* approach (e.g., in Utilities 8, 9, 10). Most utilities are diverse, with both very small DMAs (300 connections), as well as very large ones (more than 3000). The median point is around 1000 connections. Based on our experience, we recommend 500-2,000 connections (median: ~1,000) as the ideal range.

Figure 1. Distribution of the number of connections per DMA, for different water utilities.
(note the logarithmic scale)




Can you spot your utility on this spectrum? Next, let's look at the total length of the distribution network in each DMA, in Fig. 2.

Figure 2. Distribution of the total pipe lengths per DMA, for different water utilities.
(note the logarithmic scale)

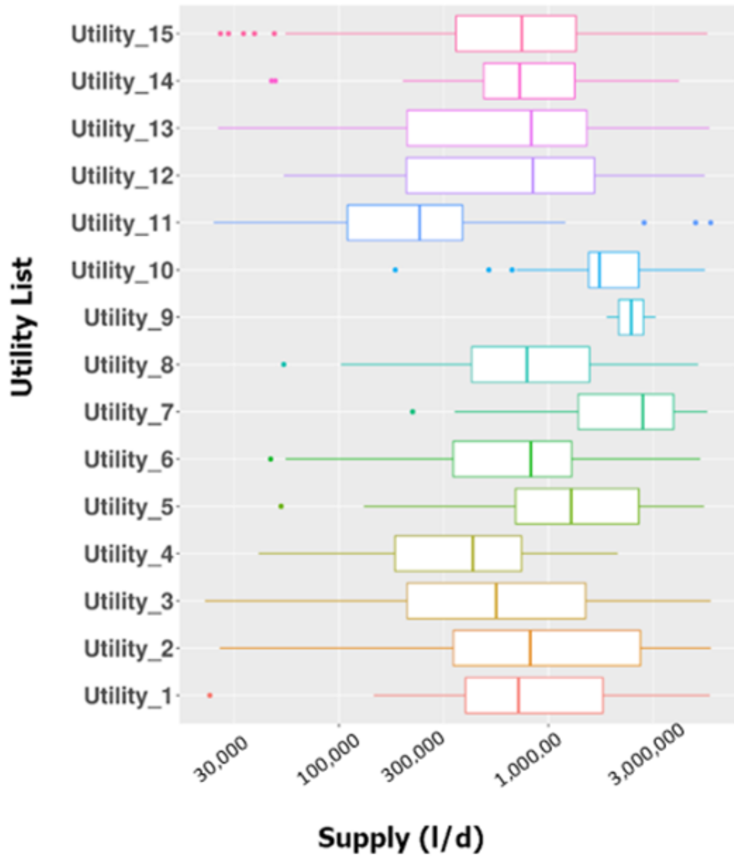



*Swarms* and *fat bulky DMAs* are visible in pipe lengths as well. The most common lengths are between 10 km and 100 km, with a median of just over 15 km. Based on our experience, we recommend 10 km - 40 km as the ideal range.

This pipe length metric is not 100% correlated with the number of connections shown in Fig. 1. At the extreme small end, Utility 3 has both a low number of connections and short pipe length, and at the extreme large end Utilities 9 and 10 have both a large number of connections and a long pipe-length metric. However, other Utilities are not consistently extreme (e.g., Utilities 8 and 11).

So, which is the best metric to use in planning DMA sizes? DMAs are all about measuring the supply. So the best option is to directly measure your DMA size in terms of supplied daily volume (Fig. 3). If you are simply making changes to your current DMA structure, that's a viable approach. But, during the DMA planning stage, this approach is mostly unavailable, so you must choose between measuring DMA size according to number of connections or according to pipe length. In that case, since it is possible to have a very long pipe, with no connections and no supply in it at all, you should measure your DMA using the number of connections, which better represents supply.

Figure 3. The distribution of daily volume supply to the DMA, for different water utilities.
(note the logarithmic scale)




The typical range is 100,000 to 3,000,000 liters/day (100 - 3,000 $m^3$/day), with a median of about 720,000 liters/day. As we will later see the square root of the supply is related to the detectable leak sizes. Mostly, the daily supply distributions are consistent with the ones found by the other metrics. More on this, in the next section.

## A Recipe for Indirect DMA Size Measurements

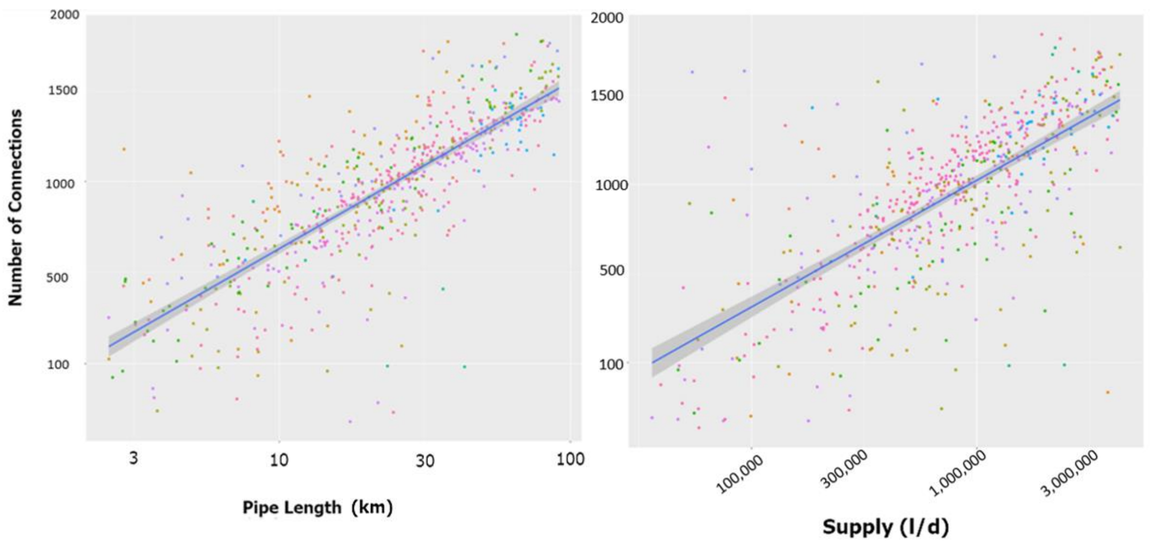
We know that DMA size is related to the total pipe length, the total daily supply, and the number of connections. But how can we calculate one metric based on the other two?

The relationships between our three DMA size metrics are determined, like most other distributions related to complex networks, by a power law. The story on power laws and their statistical relations to networks is a long one, but the law is extremely simple: $C = m \cdot L^{\alpha}$.

For this calculation, **C** is the number of connections, **L** is the sum of distribution pipe lengths, and **S** is the daily supply.

Fig. 4 depicts the power law relationships between **C**, **L** and **S**:

Figure 4. Power law relationships between the number of connections, the total pipe length, and the supply of each DMA. Different colors represent different water utilities.




The correlations shown in Fig. 4 between **C** and **L**, and between **C** and **S** are not perfectly equal (the correlation coefficient between C and L is 0.83 and between C and S is 0.67), but they provide a good recipe for calculating C, L and S based on each other:

$C = 0.22 \cdot L^{0.83}$ ; $C = 0.1 \cdot S^{0.67}$ (where S is in liters/day, and L is in m)

With actual numbers, our recipe looks like this: If your DMA total pipe length L equals 10,000 m, your connection count is probably around 459 connections (due to 0.22*10000^0.83). If you have a total daily supply of 1,000,000 liters/day, your connection count is probably around 1047 (0.1*1000000^0.67).

In the next section, we will show how this recipe is useful.

## Smallest Detectable Leak Size

How do we detect leaks in a DMA? We look for *changes* in the data, compared to the anticipated values calculated by the system – i.e., predictions. This provides a baseline pattern. When it is surpassed by a *significant amount,* you have a telltale sign of an anomaly.

How large is a “significant amount”? It depends on the normal variability of your data (*normal* means not during a leak). If it varies from minute to minute by, say, 1 liter/sec, an 0.1 liter/sec leak is much smaller than the existing fluctuation in the data, and hence impossible to detect. So, there you have it: there is a *smallest detectable* change that could be attributed to leaks.

But, what is this *smallest detectable leak*? We need some math to answer this question and you can find the details in Appendix A. Shortcutting to the practical recipe, it is given in the following equation:

**Smallest detectable leak size** = $\chi \cdot 5.62 \cdot Connections^{0.75}\ [liters/day]$

Where $\chi$is an empirical number, which we recommend to take as 10.

In a DMA with 300 connections, the smallest detectable leak size is 4,054 liters/day (or 0.05 liters/sec), and in a DMA with 4,000 connections the smallest detectable leak size is 28,284 liters/day (or 0.32 liters/sec). Note that our recipe sets a lower limit. Meter and logger inaccuracies and malfunctions, unmonitored large consumers, and other issues are likely to set the bar higher. A few examples follow.

Figure 5: An example of a leak that was detected in this DMA with 200 connections. The size of the leak is 0.33 l/s.


*Figure 5.1 - predicted behavior (green line) vs. actual behavior (blue line)*


*Figure 5.2 - night flow*

## Too Small to be True

So far, we have focused on the upper end of DMA sizes. Large DMA sizes imply that we will not be able to detect small leak events. But what about the too-small DMAs?

In Figs. 1-3 we presented the distribution of various DMA sizes, all are operational and enable leak detection. However, not all of them are equally efficient. The supply patterns of some DMAs are harder to predict (and hence, limit leak detection options), and some are easy. To quantify this, we will use empiric values of the supply data quality.

TaKaDu's quality benchmark includes two components: (1) A score (between 0 and 100) for how jittery the supply is. That is, how much it fluctuates from minute to minute; (2) a score (0 to 100) for how periodic the supply is. If the data is smooth and it also repeats itself day to day and week to week, it is easier to detect small changes when they occur.

Figure 6: quality scores (jitter and periodicity) vs. the number of connections.

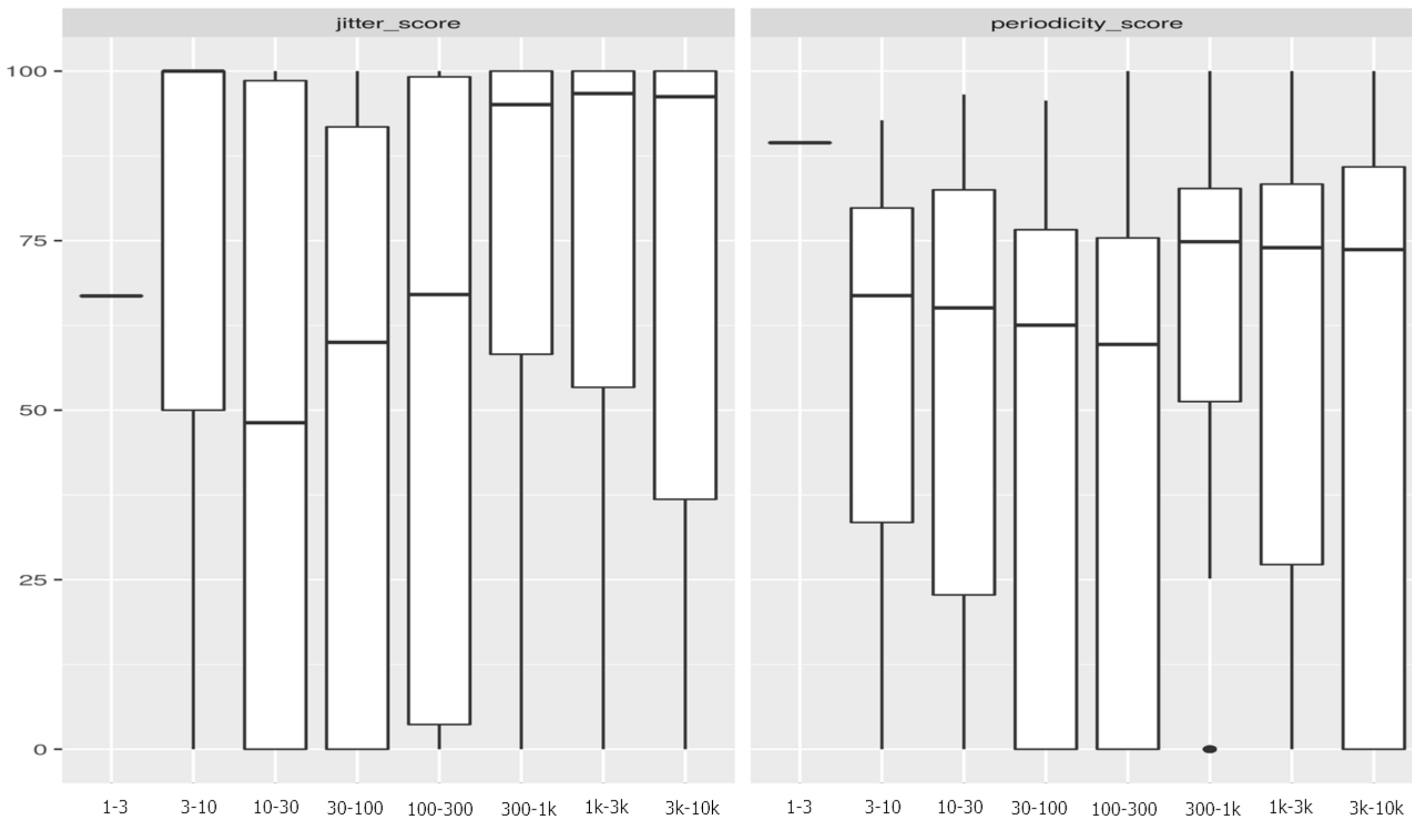



Fig. 6 shows that both the jitter score and the fluctuation score decline at around 100-300 connections ($10^2 - 10^{2.5}$ ). Note that the medians (the horizontal line inside each box) clearly decline at below 300 connections, while staying very stable above this number. However, the supply data quality is probably related to more than just the DMA size, and the spread of values within each group of DMA sizes is large. DMA sizes of $10^{2.5} - 10^3$ (300-1000) connections yield the most robust choice.

## Conclusion

You now have concrete numbers to plug in when deciding on your DMA architecture. You can compare yourself to 15 other water utilities, utilize a simple formula to know what leaks you will be able to ideally find in your planned DMA, and make sure that your data quality is not compromised by choosing a DMA that is too small.

We hope that you found this article helpful. For more information, please contact us. *info@takadu.com*

## Appendix A – The smallest detectable leak

Let's begin from the supply. The average consumer uses a volume of V every day. The total supply is $S = N \times V$ (N is the number of consumers). The standard deviation of V, $\sigma_V$, measures the variability we expect from a single consumer across many days.

The central limit theorem for sums states that the total supply standard deviation is:

$\sigma_S = \sqrt{N} \times \sigma_V = \sqrt{S} \times \frac{\sigma_V}{\sqrt{V}}$. Thus, the natural statistical variability of our supply measurement from one measure point to the next one is proportional to the square root of the total supply, with the proportionality coefficient $\chi = \frac{\sigma_V}{\sqrt{V}}$. This coefficient is not known, but its order of magnitude is 1. In practice, TaKaDu has found empirically (using our own models for leak detection) that it is more practical to expect $\chi = 10$.

Based on the above, the smallest detectable leak is around $\chi \cdot \sqrt{Supply}$. Since the supply is many times not known during the DMA planning time, a more useful formula will use the number of connections. Using our power laws from the previous section to calculate the square root of the supply, we obtain:

**Smallest detectable leak size** = $\chi \cdot \sqrt{(10 \cdot Connections)^{1.5}} = \chi \cdot 5.62 \cdot Connections^{0.75}\ [liters/day]$

As mentioned, we recommend using $\chi = 10$.