Total Site Integration for Non-Continuous Sites: Leveraging Machine Learning & Mathematical Programming

Timothy Gordon Walmsley,a\* Keegan K. Hall,a Jasper V.M. Walden,b Andreja Nemet c

aAhuora – Centre for Smart Energy Systems, School of Engineering, University of Waikato, Hamilton 3240, New Zealand

bGerman Aerospace Center (DLR), Institute of Low-Carbon Industrial Processes, Simulation and Virtual Design, Walther-Pauer-Straße 5, Cottbus, 03046, Brandenburg, Germany

cFaculty of Chemistry and Chemical Engineering, University of Maribor, Smetanova ulica 17, 2000 Maribor, Slovenia

tim.walmsley@waikato.ac.nz

Abstract

This study explores the enhancement of Total Site Integration (TSI) for non-continuous industrial sites. Utilising a combination of machine learning techniques and mathematical programming, the research innovatively addresses the challenges of defining zones for practical heat recovery. Spatial and temporal clustering methodologies are employed to achieve a more nuanced approach to TSI. The results demonstrate the significant shift in heat recovery and utility targets, highlighting the importance of zoning and the potential of integrating data-driven techniques with engineering principles in TSI practices. Depending on the temporal cluster acceptability, TSI zoning changed the utility targets by 41 % - 143 %, which underscores the criticality of the zoning approach.

**Keywords**: process integration, non-continuous processing, pinch analysis, total site integration, machine learning, mathematical programming.

* 1. Introduction

Total site integration (TSI) is a well-known methodology to optimise heat recovery and utility systems across large and multi-plant industrial sites (Klemeš et al., 1997). At its heart, TSI divides a site into discrete zones. The approach then prioritises process-to-process heat recovery within each zone before subsequently capitalising on the utility system to mediate heat recovery inter-zone. Notably, TSI is not confined to the site boundaries (Perry et al., 2008); it can equally extend to potential power, heating, and cooling exchange (Lee et al., 2020) with proximate renewable energy generation, industries, districts, and other community needs (e.g., water desalination).

Dividing a site into zones is non-trivial for many non-continuous industries where the goal is to produce practical heat integration targets. Non-continuous sites must grapple with their additional intricacies when attempting to apply TSI in practice (Tarighaleslami et al., 2017). Variables such as fluctuating stream temperatures and flow rates, flexibility versus inflexibility of target temperatures, and reliance on hot water utility add layers of complexity. To be viable, TSI strategies for these sites must be conservative to always guarantee operational integrity. Zone segmentation is pivotal in TSI but its praxis is rarely discussed in detail.

This study delves into the adaptation of TSI for non-continuous sites, specifically analysing an industrial hot water system, aided by machine learning techniques and mathematical programming.

* 1. The challenge of defining zones for total site analysis

Defining zones are an essential part of the praxis of TSI. Zones in TSI refer to defined groups of process streams that are distinct thermal regions within a site. Each zone is characterised by its temperature range and the processes operating within that range. This division helps in identifying the energy deficits and surpluses of different areas within the site. Most studies choose zones based on the operational plant units, applying the logic that intra-plant streams are close together and operationally synchronous. However, in practice, this is only sometimes true and does not assure that the resultant targets and networks meet TSI goals and operational mandates.

* + 1. Spatial clustering considerations

Spatial clustering aims to group streams that are located close to each other to avoid the complexities and costs associated with excessive and intricate heat exchanger and piping networks. Spatial locations are three-dimensional positions within a site. Proximity relates to all three coordinates being within a reasonable distance; however, there is no one-size-fits-all and heavily depends on the application. The proximity of sources and sinks is crucial for a system design that is simple, economical, and controllable.

It is common for some streams to spatially exist in two or more zones. For example, a process flow might originate in one zone and then be transported to another for further processing. In such cases, the process flow may require heating or cooling, which could happen in either zone. A decision must be made regarding the heat load allocation between the two zones, which defines the stream temperature at which it crosses the zonal boundary. This temperature can be manipulated to maximise overall heat recovery.

Appropriate spatial clustering often relies on good process knowledge and engineering judgment. A potentially complementary approach is to apply a machine learning approach, such as K-means clustering. Such clustering techniques are data-driven and require specification of the number of clusters (i.e., zones), which is often determined by heuristics, prior contextual information or arbitrary values. As a result, machine learning results provide insight but not definitive answers. Good engineering remains essential.

* + 1. Temporal clustering considerations

Most TSI studies assume that streams within the same processing plant are both spatially and temporally aligned. For some industries, this holds; but, not for all. Temporal clustering considerations focus on how well zonal source and sink heat loads match over time. Again, engineers have a significant sway in the praxis of TSI. Solely relying on judgment and experience to decide temporal compatibility, however, can lead to continued inaction.

Machine learning approaches that can support better decision-making are correlation matrix and hierarchical clustering (Müllner, 2011). Both procedures rely on determining the “distance” between each pair of time-series datasets. Distance is often a measure of dissimilarity where a distance of zero means the two series are identical. A challenge with using plant data is that it may involve similar movement patterns with short time-delays causing them to be slightly out of phase. As a result, techniques such as dynamic time warping (DTW), which is applied in this study, allow for elastic transformations of the time dimension, making it suitable for typical plant data. Using the distances from DTW analysis, hierarchical clustering provides a vantage point of possible temporal clusters depending on an acceptable distance.

* + 1. What, therefore, defines a zone in practice?

A zone is a group of streams that are spatially and temporally compatible to allow direct heat integration and retrofit. In praxis for some applications, this will lead to the creation of many additional zones, which will consequentially lower heat integration targets towards more achievable and believe levels.

In defining zones, it is also crucial to distinguish between process streams and utility. Heat integration targets are based on process streams only. However, at times, the identity of a stream can be unclear, especially process water streams that are integrated like a utility. This study makes the distinction between process streams and utility by asking whether the stream flowrate may be manipulated as part of the heat integration network. If the flowrate is determined by a processing unit, it is considered fixed from the perspective of the heat integration network and, therefore, is a process stream.



Figure 1. Conceptual views of a) zones where two streams originate in one zone and end in another and b) zoning that meets the requirements of both spatial and temporal compatibility.

* 1. Methods

The overarching method combines elements of an advanced zonal pinch analysis, total site, and machine learning. Due to space constraints, the presentation of the method is kept to an absolute minimum.

**Step 1 – Machine learning for data-driven zoning:** Zoning requires a bi-level clustering approach. First, k-means clustering of geolocations provides spatial areas. Second, time-series historical plant data of temperature and volumetric flow rate provide the basis for temporal clustering. However, such measurement data are often incomplete with numerous gaps. This study interpolates to fill in the missing data using a standard third order spline approach from the Pandas library in Python. Given the completer dataset, the required heating duty is calculated, providing the basis for hierarchical clustering using a basic DTW algorithm implemented in Python. Finally, the SciPy library in Python interprets the DTW correlation matrix to perform the hierarchical clustering and draw the dendrogram. By selecting different maximum “distance” (i.e., the measurement of data dissimilarity), various temporal zones are defined and located within the already defined spatial areas.

**Step 2 –Zonal and TSI targeting with cross-zone streams:** A bi-level optimisation approach is applied where (1) zonal targets are resolved using a modified LP temperature-interval transshipment model (Papoulias and Grossmann, 1983) within (2) a total site optimisation level where cross-zone stream temperatures, splits and an intermediate recovery loop with its associated temperatures are key variables that influence the structure and solution of zonal targets. The transshipment model is implemented in Python using the library GEKKO (Beal et al., 2018) while the outer-level optimisation uses the SciPy library. The objective is to minimise utility use given the constraint of a minimum approach temperature (10°C in this study).

* 1. Case study

The case study is based on a section of a large non-continuous processing site. Due to confidentiality, the site is not identified. Historical plant data from three spatial areas and their streams form the basis of the case study (Table 1). The median, 90th and 10th percentiles of each data are presented to provide a sense of the variability of the site.

Table 1. Stream data based on 50 days of data at 5-minute intervals. Water grade G3 is the most contaminated and G1 indicates the least contaminated.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Stream** | **Spatial**  **area** | **Water**  **grade** | **Variable** | **Unit** | **Median (P50)** | **P90** | **P10** |
| H1 | A2 to A3 | G3 | Tsupply | °C | 73.0 | 74.4 | 70.9 |
|  |  |  | Ttarget | °C | 12.0 | - | - |
|  |  |  | mcp | kW/°C | 511.1 | 537.6 | 474.0 |
| H2 | A3 | N/A | Tsupply | °C | 116.3 | 120.0 | 110.8 |
|  |  |  | Ttarget | °C | 92.0 | 92.7 | 91.1 |
|  |  |  | mcp | kW/°C | 333.6 | 378.8 | 278.0 |
| H3 | A3 | N/A | Tsupply | °C | 81.9 | 84.9 | 79.9 |
|  |  |  | Ttarget | °C | 66.5 | 72.1 | 63.8 |
|  |  |  | mcp | kW/°C | 353.9 | 377.5 | 265.2 |
| H4 | A1 | N/A | Tcond | C | 75.7 | 77.5 | 73.4 |
|  |  |  | Q | MW | 39.6 | 41.8 | 34.8 |
| C1 | A1 | G1 | Tsupply | °C | 14.9 | 16.1 | 14.3 |
|  |  |  | Ttarget | °C | 150.0 | - | - |
|  |  |  | mcp | kW/°C | 243.3 | 304.9 | 243.3 |
| C2 | A2 | G3 | Tsupply | °C | 12.0 | - | 12.0 |
|  |  |  | Ttarget | °C | 73.0 | 74.4 | 70.9 |
|  |  |  | mcp | kW/°C | 203.5 | 226.0 | 166.6 |
| C3 | A1 or A2 | G3 | Tsupply | °C | 12.0 | - | - |
|  | to A3 |  | Ttarget | °C | 75.0 | - | 75.0 |
|  |  |  | mcp | kW/°C | 1041.5 | 1138.3 | 846.0 |
| C4 | A3 | G2 | Tsupply | °C | 12.0 | - | - |
|  |  |  | Ttarget | °C | 62.1 | 68.9 | 54.2 |
|  |  |  | mcp | kW/°C | 27.9 | 28.1 | 27.7 |

* 1. Results and discussion
     1. Zonal clusters of streams

Both spatial and temporal compatibility are essential for direct intra-process integration to be efficient, economical, and practical. Figure 2 presents the results of a data-driven zoning strategy for TSI praxis.

Figure 2a delineates the temporal dissimilarity (or 'distance') between various streams within the site, with the vertical axis quantifying the DTW distance and the horizontal axis enumerating the streams (H for hot, C for cold). These streams are clustered such that the height at which streams join together in the hierarchical clustering (measured by DTW distance) indicates the order of temporal compatibility. Lower connections imply a greater temporal similarity. The horizontal dash lines are various acceptable distances that will be carried over to determine the zones with both spatial and temporal compatibility, providing a critical piece of information to the TSI targeting model. In Figure 2b, the clusters are demarcated both spatially and temporally. Each zone is characterised by a fusion of temporal clustering with spatial areas. With three spatial areas and two temporal clusters (line A), there is a maximum of 6 zones; however, streams are present in only 5 of the zones.

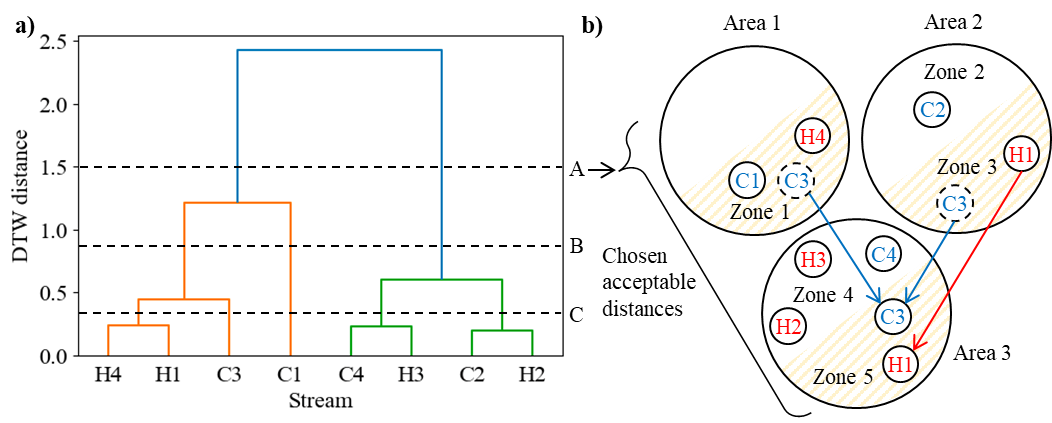


Figure 2. Zoning approach where a) the hierarchical clusters are based on DTW matrix results, showing an example acceptable distance resulting in two temporal zones and b) the spatial areas are each divided into two temporal regions (i.e., six zones) with resulting streams.

The proposed zoning method has been prototyped in Python. It involved 8 streams, resulting in the comparison of 28 possible stream pairs for 14400 time-series entries. On an M2 Macbook Air (8 GB ram), the clustering analysis took 43.5 minutes. Scaling the approach to significantly larger problems will require improved algorithmic and computational efficiency gains. One obvious inefficiency is that it currently analyses spatially incompatible stream pairs. This could be easily avoided in future implementations. Apart from computational efficiency, another challenge is the appropriate interpretation of DTW distance in deciding temporal clusters and how it relates to heat recovery between two streams. As a result, this study analyses multiple possible cluster options. Future work will look to develop a zoning heuristic.

* + 1. Total site integration targets

Maximising zonal and total site heat recovery is essential to lowering energy use and saving energy costs. However, targets need to be realistic and achievable in practice to be useful. A key to practical targets is appropriate zoning. Table 2 presents the core results from the bi-level optimisation based on different levels of acceptable temporal dissimilarity. As a result, cases A, B, and C represent differing levels of temporal and duty alignment. Cases A has 5 zones, B has 6 zones, C has 8 zones. The ‘spatial only’ serves as a baseline representing TSI performed by an engineer familiar with the plant layout without the explicit optimisation of cross zonal streams and intermediate loops as has been done historically.

Inter-plant heat recovery through the hot water utility system is substantial for this site. Tightening the acceptable temporal dissimilarity of stream-pairs leads to dividing the problem into more zones. More zones, each with less streams, results in less intra-plant heat recovery. However, the reduction in intra-zone heat recovery also gives rise to more opportunities to inter-plant heat recovery. Maximising inter-plant heat recovery requires optimising the temperatures of the hot water utility systems. As part of the outer optimisation layer, an attempt is made to manipulate the upper and lower hot water temperatures to minimise the demand for utility.

Table 2. Total site integration targets based on various temporal “distance” acceptance thresholds.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Zoning basis | Number of zones | Hot utility (MW) | Cold utility (MW) | Intra-plant heat recovery (MW) | Inter-plant heat recovery loop  (MW) |
| Spatial only | 3 | 29.4 | 1.5 | 50.8 | 32.0 |
| A | 5 | 41.6 | 13.7 | 45.1 | 25.5 |
| B | 6 | 48.7 | 20.8 | 27.6 | 35.9 |
| D | 9 | 71.3 | 43.3 | 1.4 | 39.6 |

* 1. Conclusions

Incorporating both spatial and temporal data facilitates a sophisticated approach to TSI, mindful of variations across time and space. This could lead to enhanced praxis and uptake of TSI for non-continuous industries. Although targets are gene rally higher as the number of zones increase, their practical achievability and, therefore, acceptability are improved. Future work will look to scale the approach to larger problems, improvement in computational efficiency, and heuristics to better translate the results of the clustering to zoning.

References

L. Beal, D. Hill, R. Martin, J. Hedengren, 2018, GEKKO Optimization Suite, Processes, 6(8), 106.

J.J. Klemeš, V.R. Dhole, K. Raissi, S.J. Perry, L. Puigjaner, 1997, Targeting and design methodology for reduction of fuel, power and CO2 on Total Sites, Applied Thermal Engineering, 17(8–10), 993–1003.

P.Y. Lee, P.Y. Liew, T.G. Walmsley, S.R. Wan Alwi, J.J. Klemeš, 2020, Total Site Heat and Power Integration for Locally Integrated Energy Sectors, Energy, 204, 117959.

D. Müllner, 2011, Modern hierarchical, agglomerative clustering algorithms, .

S.A. Papoulias, I.E. Grossmann, 1983, A structural optimization approach in process synthesis—II, Computers & Chemical Engineering, 7(6), 707–721.

S. Perry, J.J. Klemeš, I. Bulatov, 2008, Integrating waste and renewable energy to reduce the carbon footprint of locally integrated energy sectors, Energy, 33(10), 1489–1497.

A.H. Tarighaleslami, T.G. Walmsley, M.J. Atkins, M.R.W. Walmsley, P.Y. Liew, J.R. Neale, 2017, A Unified Total Site Heat Integration targeting method for isothermal and non-isothermal utilities, Energy, 119, 10–25.