

# Combinatorial Algorithms and High Performance Implementations for Elucidating Complex Ecosystem Relationships from North Sea Historical Data

M. A. Langston<sup>1</sup>, A. D. Perkins<sup>1</sup>, D. J. Beare<sup>2</sup>, R. W. Gauldie<sup>3</sup>,  
P. J. Kershaw<sup>4</sup>, J. B. Reid<sup>5</sup>, K. Winpenny<sup>4</sup> and A. J. Kenny<sup>4</sup>

## **Abstract**

This investigation centers on elucidating complex relationships among quantifiable variables of significance to the North Sea ecosystem. These variables encompass a huge variety of biotic and abiotic factors, and tend to possess divergent periodicities and other diverse properties. Novel mathematical tools and powerful graph algorithms are described that can be harnessed to uncover temporal, spatial and other meaningful relationships on an immense scale. High performance parallel implementations can be synthesized to extract and highlight variable sets common to multiple relationships (cliques), and to determine inflection points, putative regime changes and other patterns of possible interest. These approaches are discussed in the context of more traditional clustering methods. Data quality and the significance of missing or corrupted values are also addressed, as is the importance of mining data at multiple levels of granularity. A long-term goal is to establish data dependencies upon which we can draw conclusions about the impact of man and other agents upon the Sea.

## **1. Background**

Effective multivariate data clustering lies at the heart of attempts to understand relationships embedded in ecosystems as complex as those found in the North Sea. A central goal is to find algorithms that can extract subsets of variables with approximately similar observational profiles.

A huge variety of clustering approaches has been proposed. Unfortunately, most are limited by the fact that the clusters they produce are disjoint, requiring that a variable be assigned to a single cluster. Of course this greatly simplifies the analysis. Unfortunately, it also presupposes a model with at best only modest fidelity to the system under study. In the case of North Sea data, for example, many variables appear to play important roles in multiple but distinct relationships. Meaningful clusters should therefore overlap at these variables.

---

<sup>1</sup> Department of Computer Science, University of Tennessee, Knoxville, TN 37996-3450, USA

<sup>2</sup> Agriculture and Fisheries Unit, Institute for the Protection and Security of the Citizen, European Commission Joint Research Centre, I-21020 Ispra (VA), Italy

<sup>3</sup> Merit (NZ) Ltd., 76 Inglis Street, Wellington 6022, New Zealand

<sup>4</sup> CEFAS Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

<sup>5</sup> Joint Nature Conservation Committee, 7 Thistle Place, Aberdeen AB10 1UZ, Scotland, UK

Popular clustering approaches are also limited in that they do not recognize negative correlations. Yet negative correlations are widely witnessed, and can be particularly meaningful when viewed from an ecosystems perspective. Relevance networks have been proposed in an effort to represent complex relationships with correlation structures and thus to overcome the limitations of traditional clustering methods. Unless there is an algorithmic means to extract tightly-connected correlate subsets, however, many of the most interesting relationships will remain embedded within a vast space of correlations.

## 2. Graph Theoretical Methods

There are many benefits to the mathematical power and abstraction of a graph theoretical approach. We embrace the spirit of relevance networks, and begin with a symmetric correlation matrix,  $M$ , in which the rows and columns represent variables, and in which the entry at location  $M_{i,j}$  denotes the correlation coefficient between variables  $i$  and  $j$ . From  $M$  we build a weighted correlation graph,  $C$ , whose vertices represent variables and whose edges are annotated with correlation coefficients. With the use of a high-pass filter and a chosen threshold,  $t$ , we produce from  $C$  a simple, unweighted graph,  $G$ , whose structural properties we seek to comprehend.

Once this graph is created, the focus moves to extracting its densest subgraphs. These are tightly-coupled sections of the graph whose vertices represent strongly correlated variable subsets. Ideally, every pair of vertices in such a subgraph is connected by an edge. In this case the subgraph is called a *clique*. A clique on five vertices is illustrated in **Figure 1**. It is particularly noteworthy that cliques need not be disjoint. A vertex can reside in more than one clique, just as a variable may be in more than one relationship. Moreover, negative correlations are easily handled in a variety of ways, for example, by two-coloring the graph's edges prior to thresholding. In terms of correlation density, clique represents the most trusted potential for identifying sets of interacting variables.

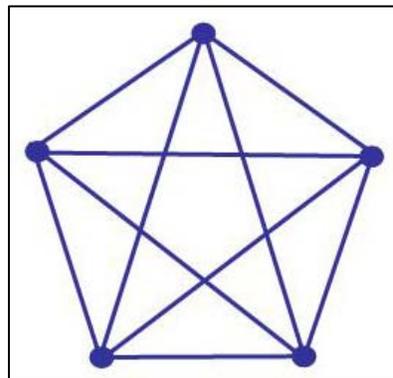


Figure 1. A clique of size five.

False positives therefore, the bane of so many analysis techniques, are seldom problematic for clique-centric methods. In fact, we must guard against false negatives, so much so that the use of Bonferroni corrections, false discovery rates and the like are generally avoided. After all, a clique of size  $k$  requires  $k(k-1)/2$  significant correlations. The loss of a single correlation means the loss of the clique. In this way clique is complementary to more traditional clustering tools. The tradeoff is that clique is extremely difficult to solve.

## 3. Pioneering Advances in the Search for Cliques

The inputs to the clique problem are an undirected graph  $G$  of order  $n$  and a parameter  $k \leq n$ . The question asked is whether  $G$  contains a subgraph isomorphic to  $K_k$ . Clique is  $NP$ -complete, and hence without any known algorithm that runs in time polynomial in

$n$  and  $k$ . Clique cannot even be approximated in polynomial time to within  $n^\epsilon$ , for any  $\epsilon > 0$ , unless  $P=NP$  [1]. Novel approaches are clearly required if clique is to be solved on data sets as huge and complex as those we consider here.

We therefore utilize *fixed-parameter tractability*, whose roots can be traced at least as far back as early work by Fellows and Langston [2, 3] on nonconstructive applications of well-quasi-order theory and the celebrated Graph Minor Theorem of Robertson and Seymour [4].

*A problem is fixed-parameter tractable if it has an algorithm that runs in  $O(f(k)n^c)$  time, where  $n$  is the problem size,  $k$  is the input parameter, and  $c$  is a constant independent of both  $n$  and  $k$ .*

The first truly thorough treatment of this subject was published by Downey and Fellows [5]. Excellent recent monographs are due to Neidermeier [6] and Flum [7].

Although clique itself is not fixed-parameter tractable, we shall solve instead the *vertex cover* problem, which is clique's complementary dual. The result is an algorithm whose run time is linear in  $n$  and whose exponential growth in  $k$  is limited to a mere *additive* term, making it realistic now to consider the search for cliques of huge sizes in immense collections of ecosystem data. For brevity, we have suppressed a great number of technical details. Our recent work on this subject is featured in [8, 9], where we deal at length with application details and put our tools to work in the context of transcriptomic data analysis.

#### **4. Scalable Implementations and High Performance Computing**

The problem described in the last section is known as the classic *decision* version of the *maximum* clique problem. It sheds light on the overall complexity of clique, and provides a basis for novel algorithms. In practice, however, we must also solve the *optimization* version of the problem, which means computing the size of a largest clique, as well as the *search* version, which means isolating a clique of that size. Thanks to problem transformations, it turns out that optimization and search are relatively easy if we have efficient methods for decision.

On the other hand, we are frequently required also to solve the *maximal* clique problem. Such a clique need not be the largest in the graph (a global measure of optimality). Instead it must only be a complete subgraph with the property that no other vertex in the graph can be added to it to complete a larger complete subgraph. Thus a maximal clique is a local measure of optimality. Because we cannot know in advance which maximal cliques are the most revealing, the problem becomes one of enumerating all maximal cliques in the graph.

Space, not just time, now becomes a critical resource. A graph with  $n$  vertices may in theory have as many as  $3^{n/3}$  maximal cliques [10]. In practice, we have found in handling many biological data sets that a maximum clique size in the teens can mean thousands of maximal cliques, while a maximum clique size nearing 100 can mean tens of millions of maximal cliques. Space requirements are compounded by the fact that algorithms for enumerating maximal cliques of a given size,  $k$ , often require maintaining all maximal cliques of size  $k-1$ .

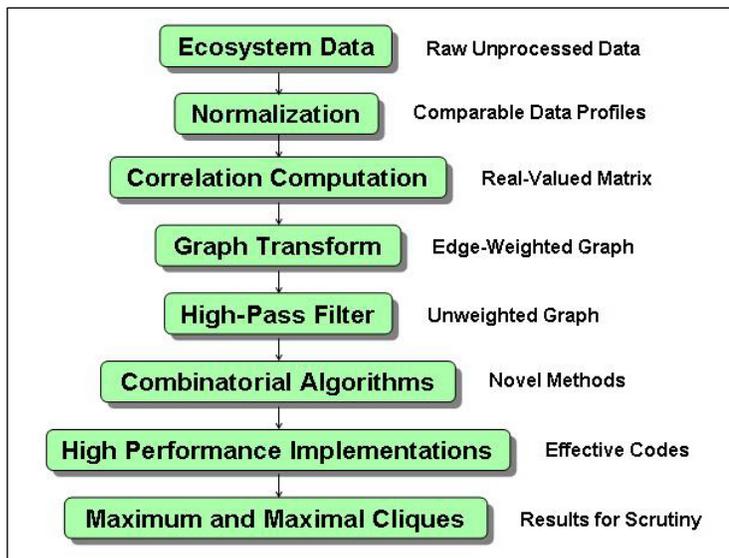
Thus, as graphs grow in size, we turn more and more to high performance platforms to solve the maximal clique problem. We have addressed the memory bottleneck just described with an algorithmic strategy that provides exact, parallel and scalable solutions. It exploits not only fixed-parameter tractability, but also ultra-large globally addressable memory architectures such as that of the SGI Altix 3700 at Oak Ridge National Laboratory (shown in **Figure 2**), as well as an innovative bitmap data representation scheme. We refer the interested reader to [11] for details.



**Figure 2. Supercomputers are often required.**

## 5. A Data Analysis Toolchain

The enormous volumes of ecosystems data now available provide a unique opportunity to investigate mathematical models of complex biotic and abiotic interactions at an unprecedented level of scale. Such models, when coupled with high performance computational means, hold the promise of predictive dynamic simulations with which we can evaluate multi-factor causality and estimate the impact of policy. In order to realize the potential of this approach, however, high dimensional historical data must first be analyzed to devise hypothetical models and determine which may be plausible and meaningful. An early step in model formulation is to determine putative relationships embedded in multivariate data. To accomplish this, we employ the algorithmic toolchain illustrated in **Figure 3**. Raw data is initially normalized to bring values into comparable ranges. Correlation coefficients are then computed for each pair of variables, for example, by using Pearson's or Spearman's rank calculations. This produces the aforementioned symmetric correlation matrix,  $M$ , from which we are able to construct graphs and extract cliques as previously discussed.



**Figure 3. A toolchain for ecosystems data analysis.**

## 6. The Challenge of High-Dimensional Multivariate Ecosystems Data

The wealth of data used in this study presents several special challenges to those who would attempt to unlock its secrets. For example, a wide variety of differing sources were used to generate measurements. Moreover, these measurements were taken at varying levels of granularity. On top of that, temporal metrics were recorded some times on a monthly basis, at other times on a quarterly basis, and at yet still other times on an annual basis. Of critical significance is the fact that huge numbers of data values are missing, incomplete or of questionable reliability. Plus there are inherent biases, for example, in the amount of effort spent on each of the squares under consideration by the ICES Regional Ecosystem Group for the North Sea (REGNS). Together these many factors tend to compound the difficult problem of combining, preprocessing, integrating and normalizing data spread over so many different types and sources. For detailed specifics on REGNS data types and completeness, see [12, 13].

## 7. Representative Results

### 7.1 Correlation Computations

By composition, multivariate/megavariate data should present us with a variety of ways to compute correlations. Historical ecosystems data are no exception. We can, for example, choose any single variable to play the role of “condition” upon which we measure and compute pairwise correlation coefficients over all other variables. The goal is to select a condition for which the resulting coefficient histogram approximates a normal distribution. In this way, we can apply conventional statistical procedures.

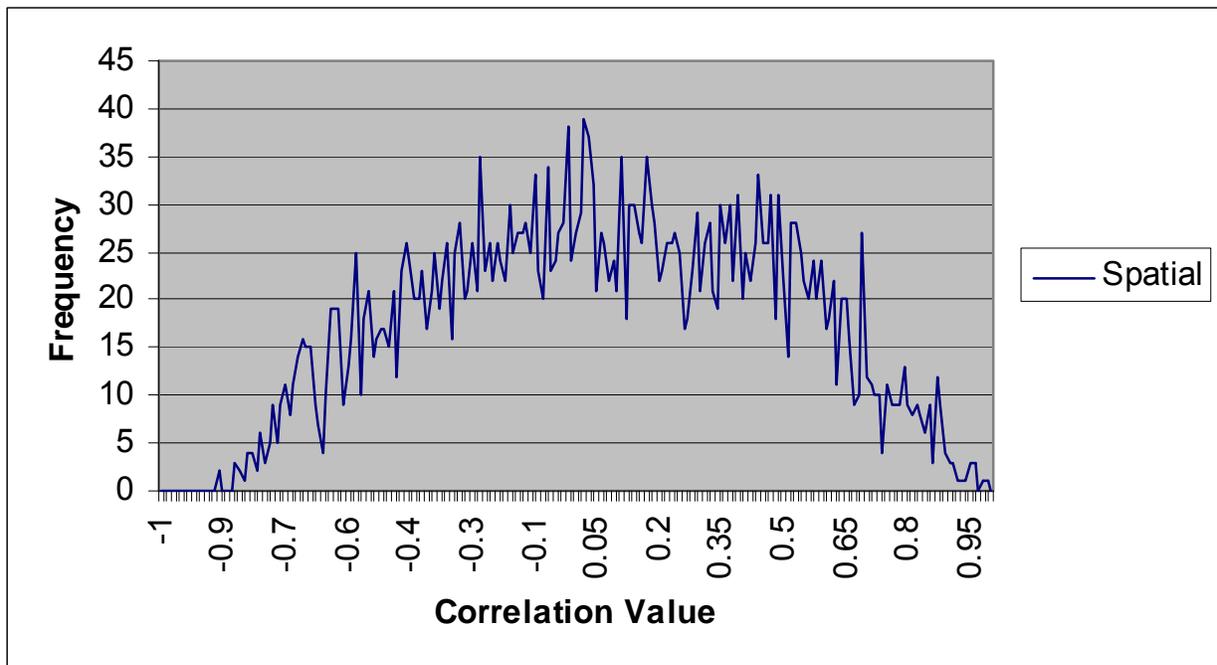
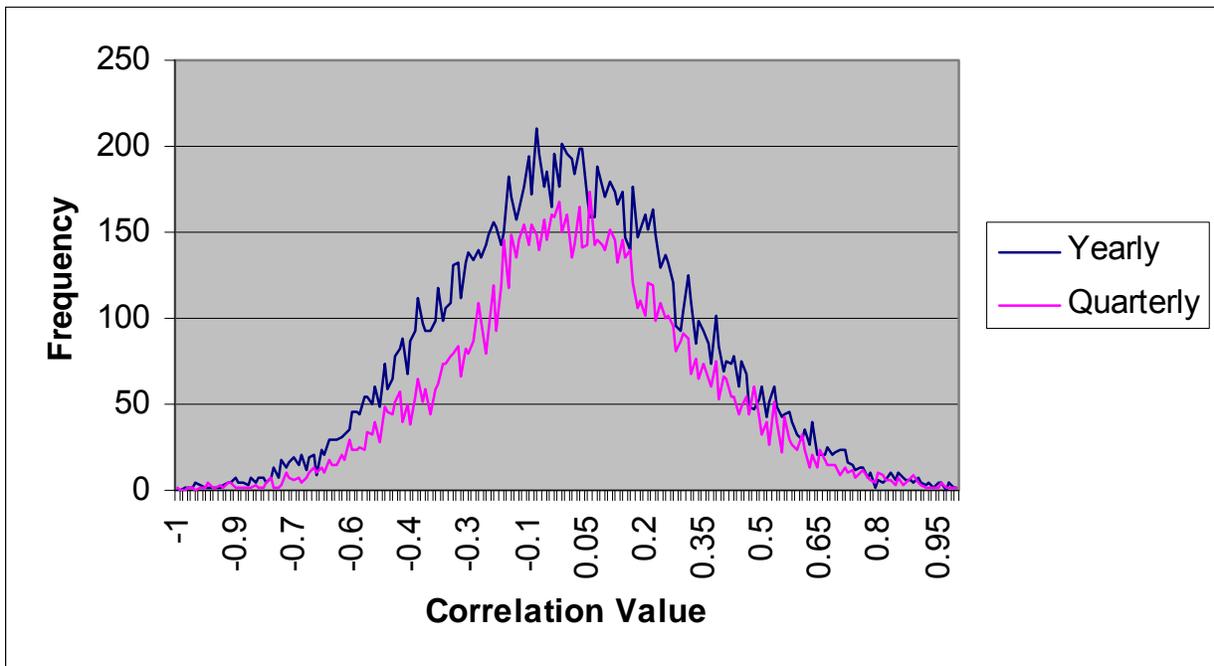


Figure 4. Histogram of correlation coefficients based upon spatial averages.

Two popular choices for the condition are space and time. In the case of North Sea data, we have found the latter generally to be superior. Debating this point is beyond the scope of this paper, however, and we merely illustrate typical histograms in **Figure 4** and **Figure 5**. Horizontal axes list correlation coefficient values ranging from -1 to +1. Vertical axes show, for each value, the number of correlations found over pairs of biotic and abiotic variables. In preparing these figures, Pearson's correlation coefficients were calculated, with those based upon fewer than twelve pairwise observations discarded. Observe the superior "tails" present when time is used to drive the correlations. These make it reasonable to try to choose a meaningful threshold with which we can obtain the unweighted graph as previously described.

There are several ways to select such a threshold. For example, practitioners have sometimes suggested 0.85 as a universal cutoff value. Other methods employ functional knowledge or match the threshold to a desired maximum clique size [14].



**Figure 5. Histogram of correlation coefficients based upon temporal averages.**

Because so many variable readings are missing in this particular dataset, however, the technique we adopt here is to compute the significance of correlations non-uniformly using the Student's *t*-test with  $N-2$  degrees of freedom, where  $N$  is taken to be the number of common observations between the two correlates. Any edge whose weight corresponds to a correlation not significant at a preselected  $\alpha$  level is removed. Weights on remaining edges are then ignored. After this we can begin the computationally intensive task of clique extraction.

## 7.2 Clique and Cluster Overlap

The discriminatory power of clique-centric methods can be illustrated with annual North Sea historical data. At  $\alpha=0.05$  (which on average corresponds to a threshold of about 0.42), we find examples such as the one shown in **Figure 6** and detailed in **Table 1**. Only positive correlations are considered, with bird species depicted in red, fish in blue, plankton in green and abiotic variables in black.

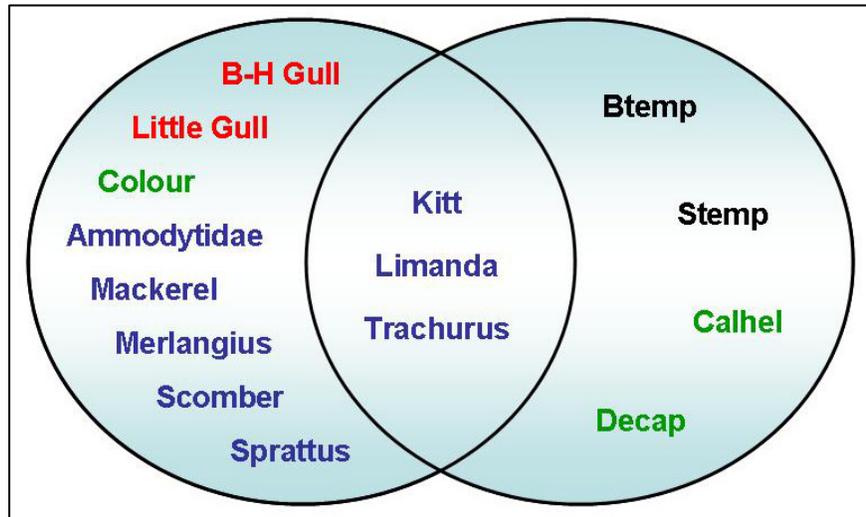


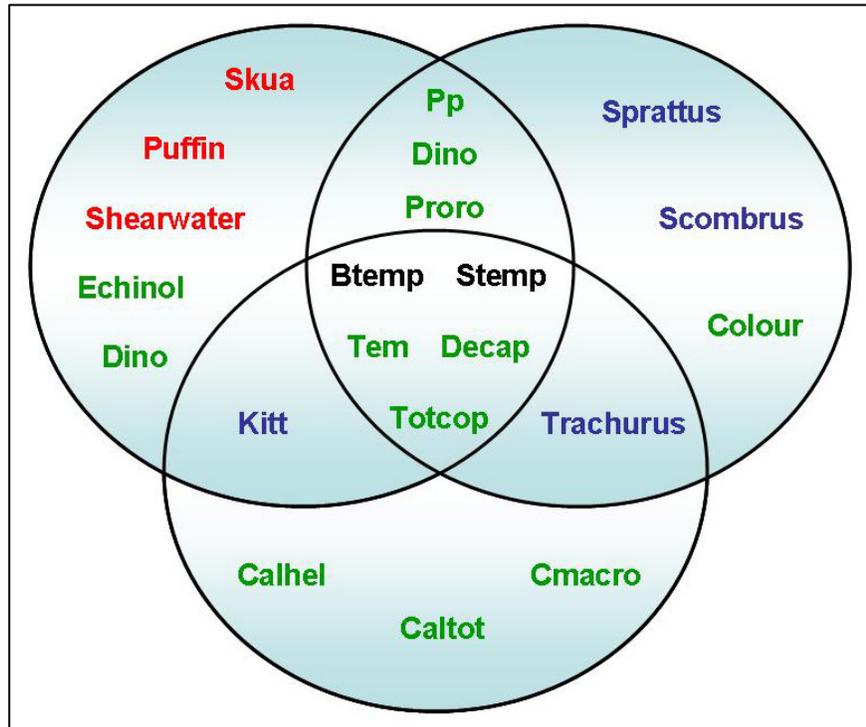
Figure 6. Cliques elucidate complex relationships.

Table 1. Details on variables shown in Figure 6.		
<b>Abiotic</b>		
Btemp	Bottom temperature	
Stemp	Surface temperature	
<b>Phytoplankton</b>		
Colour	'Greenness' from CPR screen	
<b>Zooplankton</b>		
Calhel	<i>Calanus helgolandicus</i>	
Decap	Decapoda total	
<b>Fish</b>		
Ammodytidae	<i>Ammodytidae</i>	Sand lance
Kitt	<i>Microstomus kitt</i>	Lemon sole
Limanda	<i>Limanda limanda</i>	Common dab
Merlangus	<i>Merlangius merlangus</i>	Whiting
Scombrus	<i>Scomber scombrus</i>	Mackerel
Sprattus	<i>Sprattus sprattus</i>	Sprat
Trachurus	<i>Trachurus trachurus</i>	Horse mackerel
<b>Birds</b>		
B-H Gull	<i>Larus ridibundus</i>	Black-headed Gull
Little Gull	<i>Larus minutus</i>	Little Gull

Note the significance of overlapping cliques, and the role three species of fish play at the nexus. Variables not in a common clique are poorly correlated. The correlations between Ammodytidae and the four variables Btemp, Stemp, calhel and decap, for

example, have respective p-values 0.37, 0.60, 0.97 and 0.61. Traditional clustering methods, however, would either have forced Ammodytidae and these four variables to reside in the same cluster, or have removed one or more of the three fish in the intersection from the cluster highly populated with fish or from the cluster with abiotic factors (or both). In each of these cases, the accuracy with which the resultant clusters represent the underlying ecosystem data would have been severely compromised.

By using quarterly data rather than annual, we are sometimes able to see more detailed relationships. At  $\alpha=0.01$  (which on average corresponds to a threshold of about 0.30), we discover a huge number of overlapping cliques, three of which are shown in **Figure 7** and detailed in **Table 2**. As before we consider positive correlations only, with bird species again depicted in red, fish in blue, plankton in green and abiotic variables in black. At this level of scrutiny, it becomes clear that



**Figure 7. Lower thresholds can reveal more detailed structures.**

plankton plays a major role, and that something as simple as an abiotic factor like temperature can be at the confluence linking many seemingly dissimilar variables.

### 7.3 Cluster Refinement and the Paraclique Algorithm

Noise is a constant concern when handling historical data. We have previously developed noise-reduction techniques, most notably the “paraclique” algorithm, for dealing with high-throughput biological data [15]. We apply paraclique here in order both to mitigate the effects of noise as well as to view correlation structures at a more interpretable level of granularity. After all, it can be difficult indeed to try to view several thousand overlapping cliques at one time.

Informally, a paraclique is an extremely densely-connected subgraph, but one that may be missing a small number of edges. In the present application, this corresponds to a very highly correlated group of variables whose representational levels show highly significant but not necessarily perfect pair-wise correlations. Although the complete algorithm is fairly complex, a simplified version of it may be described as follows. Beginning with a clique,  $C$ , of size  $k$ , we consider each non-clique vertex,  $v$ , in turn. We

Table 2. Details on variables shown in Figure 7.		
<b>Abiotic</b>		
Btemp	Bottom temperature	
Stemp	Surface temperature	
<b>Phytoplankton</b>		
cmacro	<i>Ceratium macroceros</i>	
Colour	'Greenness' from CPR screen	
Dino	<i>Dinophysis spp.</i>	
Proro	<i>Prorocentrum spp.</i>	
<b>Zooplankton</b>		
Calhel	<i>Calanus helgolandicus</i>	
Caltot	<i>Calanus</i> total traverse	
Decap	<i>Decapoda</i> total	
echinol	<i>Echinoderm</i> larva	
Pp	<i>Para-Pseudocalanus spp.</i>	
<b>Fish</b>		
Kitt	<i>Microstomus kitt</i>	Lemon sole
Scombrus	<i>Scomber scombrus</i>	Mackerel
Sprattus	<i>Sprattus sprattus</i>	Sprat
Trachurus	<i>Trachurus trachurus</i>	Horse mackerel
<b>Birds</b>		
Shearwater	<i>Puffinus puffinus</i>	Manx Shearwater
Skua	<i>Stercorarius parasiticus</i>	Arctic Skua
Puffin	<i>Fratercula arctica</i>	Atlantic Puffin

mark  $v$  if and only if it is adjacent to at least  $k-1$  vertices in  $C$ . After each vertex has been considered, we define a paraclique,  $P$ , to be the union of  $C$  and the set of all marked vertices. We then remove  $P$  from the graph and iterate.

Sample results of the paraclique method are illustrated in **Figure 8**, where the correlates are the 208 squares defined for the North Sea, each roughly 30 kilometers by 30 kilometers in size, with correlations driven by the average annual readings of all other variables. By harnessing the computational power of fixed-parameter tractability and then isolating paracliques, we are able to identify considerably denser subgraphs than are typically produced with traditional clustering algorithms. While we have observed edge densities ranging in the 10-20% range with simple  $k$ -cores and neighborhoods, and in the 50-60% range with hierarchical clustering schemes, on real data paraclique consistently seems to return subgraphs with densities upwards of around 95%. Of course within a paraclique all variables are highly positively correlated. Thus it can be revealing to examine negative correlations. It is noteworthy that among the paracliques shown in this figure, each of those located in the north has the property

that all of its squares are significantly negatively correlated with every square in at least one paraclique located in the south. It is also interesting to compare this distribution of paracliques with the spatial distribution of clusters of abiotic, plankton, fisheries and birds data computed with more traditional methods [12]. There is a high degree of agreement between these results, particularly between paracliques and the clusters derived for abiotic and plankton data. The derivation of similar contiguous patterns of corresponding variables, using very different algorithmic approaches, is encouraging.

## 8. Moving Forward

This paper is mainly focused on novel methods, and with new methods come many as-yet unanswered questions. For brevity we mention just one, namely, the exploration of putative regime changes in the sense of [16] in an effort to assess the impact of human and other pressures on the North Sea. In this we are not really limited by the immense computations required, but rather by our inability to isolate the most meaningful factors. Any evidence for such a regime change is apt to be highly dependent on the variables used to drive its correlations. The result is that, depending on the variables chosen, there may be multiple, conflicting claims of regime changes, including the claim there has been no change at all. If we choose to let *all* variables drive the correlations, the best we can say at this time is that two sets of years seem to indicate some sort of change in overall North Sea readings. The four years from 1984 to 1987 form one clique; the three years from 2002 to 2004 form another. Although all correlations are of course positive within each clique, it turns out that each and every year in the first clique is significantly negatively correlated with each

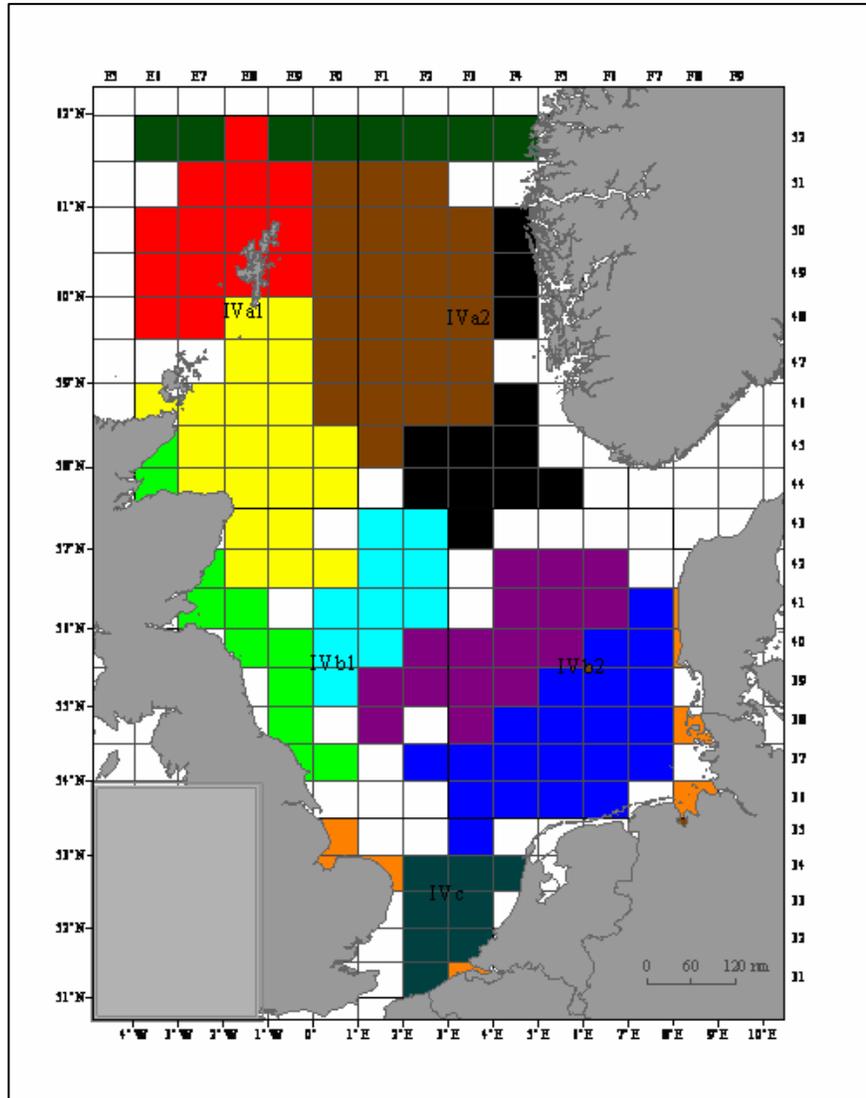


Figure 8. Paracliques of North Sea historical data.

and every year in the second clique. This is a surprisingly strong computational finding. What is to be made of it? And what is to be made of the missing years? The fact that so many years are not associated with any cliques at all suggests to us that using the entire set of variables to drive correlations has a dilutive effect, and that one should indeed be selective in the biotic and abiotic variables chosen.

## **9. Remarks**

The North Sea provides a rich and representative ecosystem for detailed study. There are many inviting opportunities for system evaluation, modeling and management. On the other hand, high dimensional multivariate ecosystem data by its very nature present a number of strikingly difficult challenges to any serious attempt to analyze the complex relationships among its biotic and abiotic factors. This situation creates the universal ecologist's problem: many variables are correlated with one another, but with statistically weak pairwise correlations. Combinatorial analysis based on the clique problem effectively compiles these weak correlations into statistically robust inter-related clusters.

We have therefore synthesized and implemented our innovative algorithms for use on high performance computational platforms. With these, we are able to mine and extract relevant system constituents in an effort to uncover important relationships and other possible patterns of interest. In particular, clique-centric analysis offers us the opportunity to identify gaps in our collective datasets where important relationships are likely to exist. Although this paper has focused primarily on novel methods, a sampling of preliminary results has been chosen to help illustrate the potential of this approach in establishing key data dependencies. In this role, clique can be viewed as a powerful generator of hypotheses amenable to further testing.

Neither fisheries biologists nor computational scientists can work in a vacuum if we are to make real progress. These and other experts must pool their collective knowledge in order to reach a new level of understanding about ecosystems that are as complex as the North Sea. We are optimistic that continuing research along this line will help us to draw far-reaching conclusions about the impact of man and other agents upon the world's oceans.

## **Acknowledgments**

We wish to express our gratitude to ICES and especially to Adi Kellermann for granting us access to the rich source of North Sea historical data used in this study. We also wish to thank all those whose efforts have helped record, compile, curate and evaluate this important resource. The work reported in this paper has been supported in part by a seed money research grant from the University of Tennessee.

## Contact Information

Correspondence concerning this article should be addressed to Professor Michael A. Langston, Department of Computer Science, 203 Claxton Complex, 1122 Volunteer Blvd, The University of Tennessee, Knoxville, TN 37996-3450, USA. Email: [langston@cs.utk.edu](mailto:langston@cs.utk.edu). Phone: 1-865-974-3534.

## References

- [1] U. Feige, S. Goldwasser, L. Lovasz, S. Safra, and M. Szegedy., Approximating the maximum clique is almost *NP*-complete, *Proceedings, IEEE Symposium on the Foundations of Computer Science*, 1991, 2-12.
- [2] M. R. Fellows and M. A. Langston, Nonconstructive Tools for Proving Polynomial-Time Decidability, *Journal of the ACM*, vol. 35, 1988, 727-739.
- [3] M. R. Fellows and M. A. Langston, On Search, Decision and the Efficiency of Polynomial-Time Algorithms, *Journal of Computer and Systems Sciences*, vol. 49, 1994, 769-779.
- [4] N. Robertson and P. D. Seymour, Graph Minors. XX. Wagner's conjecture, *Journal of Combinatorial Theory, Series B*, vol. 92 (2), 2004, 325-357.
- [5] R. G. Downey and M. R. Fellows, *Parameterized Complexity*. New York: Springer) 1999.
- [6] R. Niedermeier, *Invitation to Fixed-Parameter Algorithms*: Oxford University Press) 2006.
- [7] J. Flum and M. Grohe, *Parameterized Complexity Theory*: Springer) 2006.
- [8] F. N. Abu-Khzam, M. A. Langston, P. Shanbhag, and C. T. Symons, Scalable parallel algorithms for FPT problems, *Algorithmica*, vol. 45, 2006, 269-284.
- [9] F. N. Abu-Khzam, M. R. Fellows, M. A. Langston, and W. H. Suters, Crown Structures for Vertex Cover Kernelization, *Theory of Computing Systems*, 2006, accepted for publication.
- [10] E. Tomita, A. Tanaka, and H. Takahashi, The Worst-Case Time Complexity for Generating all Maximal Cliques, *Proceedings, Computing and Combinatorics Conference*, Jeju Island, Korea, 2004.
- [11] Y. Zhang, F. N. Abu-Khzam, N. E. Baldwin, E. J. Chesler, M. A. Langston, and N. F. Samatova, Genome-scale computational approaches to memory-intensive applications in systems biology, *Proceedings, Supercomputing*, Seattle, Washington, 2005.
- [12] A. J. Kenny, P. J. Kershaw, D. J. Beare, M. Devlin, J. B. Reid, P. Licandro, A. Gallego, K. Winpenny, C. Houghton, M. A. Langston, H. R. Skjoldal, and A. D. Perkins, Integrated Assessment of the North Sea to Identify the Relationship between Human Pressures and Ecosystem State Changes – Implications for Marine Management, *Proceedings, International Council for the Exploration of the Sea Annual Science Conference*, 2006.
- [13] D. J. Beare, A. J. Kenny, P. J. Kershaw, E. McKenzie, M. Devlin, J. B. Reid, P. Licandro, K. Winpenny, C. Haughton, M. A. Langston, H. R. Skjoldal, and A. J. Perkins, Building Multi-Discipline, Multivariate Databases for Use in Integrated Assessments: Experiences and Recommendations, *Proceedings, International Council for the Exploration of the Sea Annual Science Conference*, 2006.
- [14] E. J. Chesler, L. Lu, S. Shou, Y. Qu, J. Gu, J. Wang, H. C. Hsu, J. D. Mountz, N. E. Baldwin, M. A. Langston, J. B. Hogenesch, D. W. Threadgill, K. F. Manly, and R. W. Williams,

Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function, *Nature Genetics*, vol. 37 (3), 2005, 233-242.

[15] E. J. Chesler and M. A. Langston, Combinatorial genetic regulatory network analysis tools for high throughput transcriptomic data, *Proceedings, RECOMB Satellite Workshop on Systems Biology and Regulatory Genomics*, San Diego, 2005.

[16] M. Weijerman, H. Lindeboom, and A. F. Zuur, Regime shifts in marine ecosystems of the North Sea and Wadden Sea, *Marine Ecology Progress Series*, vol. 298, 2005, 21-39.