Visualising Energy Data

**Diane Donovan, Trevor Pickett and Richard Wilson*

University of Queensland, Brisbane, Australia

Abstract: The Australian energy market is in the final stages of deregulation. These changes have created an exciting and dynamic environment which is highly volatile and competitive with respect to both demand and price.

Our current research seeks to visualise aspects of the National Energy Market with a view to developing techniques which may be useful in identifying trends and/or drivers of these trends. In order to capture the complexity of the problem we have explored a suite of different visualisation techniques, which, when combined into a unified package, highlight diverse aspects of the problem at hand. The particular problem to be visualised in this manner is "Does the data exhibit characteristics which suggest that the time of day, day of the week, or the season, affect the variation in demand and/or price?"

1 Introduction

The main purpose of this study is to investigate the predictive powers of various visualisation techniques in identifying trends and/or drivers of these trends. The physical setting for this study is the National Electicity Market, and we investigate the conformance of participants to trends during varying timeframes.

The energy industry is of interest as there is a dynamic interplay between production and use, transfer and storage, and purchase and sale, and when these factors are coupled with issues of weather and technological advances they play a major role in driving the market. The complexity of these underlying forces together with vast amounts of data make it difficult to establish long term trends and thus the provision of useful visualisation requires data simplification and careful variable selection.

Much of the existing analysis of data is conducted on half hourly intervals over an entire day, and across all 365 days of the year. By utilizing various visualisation techniques we will show that the data exhibits trends which are relatively constant for certain timeframes in any given day and that there appears to exist trends which develop over a twelve monthly period. However, the analysis presented here does not reveal trends which occur over a seven day cycle and it is anticipated that further refinment of the data will be required here (such analysis is part of the ongoing work). An important aspect here is the use of different visualisation techniques to highlight varying trends within the data.

To ensure subsequent discussions are meaningful we begin the paper with a description of the relevant aspects of the Australian energy market. This is followed by a more detailed discussion of the relevant data sets. A suite of visualisation techniques have been utilised at various times and will be discussed in the latter stages of the paper.

2 Relevant Background in the Energy Market

The National Electricity Market (NEM) was established in December 1998 as part of a process to deregulate Australian energy markets. The aim was to provide for the wholesale supply and purchase of electricity in five Australian states and territories; New South Wales, Victoria, Queensland, South Australia and the Australian Capital Territory. NEM supplies energy to 7.7 million Australian customers and trades approximately 8 billion dollars of energy per year. The market is based on an open access grid consisting of transmission and distribution networks across all states and territories. It seeks to promote a competitive environment at each stage in the electricity production and supply chain.

Currently the NEM is divided into five interconnected electrical regions, which are the above five states, with the ACT incorporated into NSW and the Snowy designated as a fifth region. Each region has a regional reference node which is usually a major city or generating centre. Generally all electricity in the NEM is traded through a spot market managed by the National Electricity Market Management Company Limited (NEMMCO). This spot market provides the interface between generating operators and electricity retailers. Generating operators provide bids, comprising of price and associate quartiles, and offer to supply NEMMCO differing amounts of energy at given prices. In general, generators sell electricity into the pool, and retailer and wholesaler users pay for the electricity they draw from the pool. NEMMCO ensures that there is enough electricity to meet demand. For each reference node, in each of the five regions, NEMMCO uses the bids and constraints, such as interconnecting capacities and energy transmission and distribution losses, to set a Regional Reference Price (RRP). Based on the combined information NEMMCO centrally co-ordinates the dispatch process, and schedules generators by sending generating instructions every five minutes. Decisions are made on the principle of determining the most efficient way to meet demand.

3 The Data

NEMMCOs website [**2**] provides the general public with access to NEM data. For each region we have accessed data on Total Demand, in megawatts, and the Region Reference Price, RRP (in dollars per megawatt hour). NEMMCO provides an online repository of this data, available in 30 minute intervals from October 13, 1998. We have sought to compare the change in relative demand and price for Queensland against the change for a combination of other NEM regions.

Before visualisation could be possible the data was massaged into a workable format.

First, for each region, the Total Demand and RRP data was been stripped from the monthly files provided by NEMMCO, and combined into yearly files for each region.

The 48 half hourly timesteps were judged to be too high resolution for our requirements, so the data was averaged over two half hourly timesteps, before and on each hour of the day; that is, an average was taken of $TD_{i-0.5}$ and TD_i ($RRP_{i-0.5}$ and RRP_i), where TD_i (RRP_i) represents the Total Demand (RRP) on the hour i, for $1 \leq i \leq 24$. Table 1 illustrates this process for a small portion of the original data.

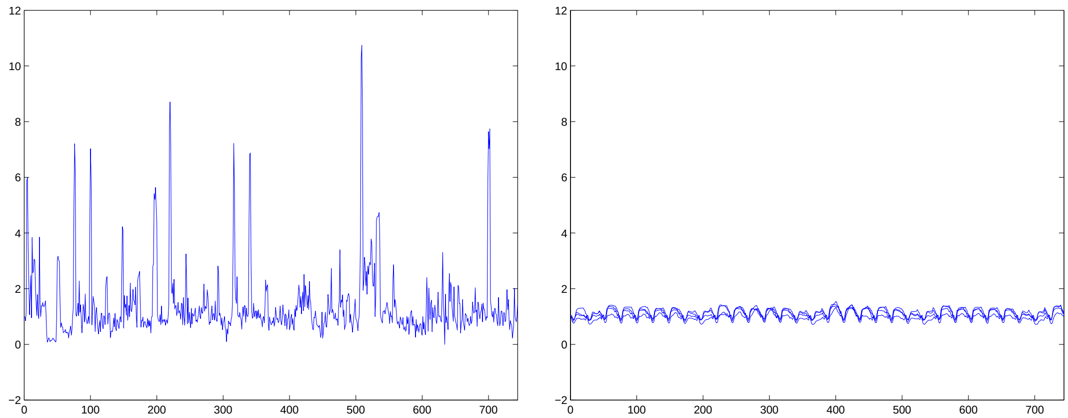
Time Step	Original Total Demand	Reduced Total Demand
2:30	4187.93167	
3:00	4090.09833	4139.015
3:30	4032.42167	
4:00	4025.80333	4029.1125
4:30	4008.265	
5:00	4025.675	4016.97
5:30	3940.10167	
6:00	3969.85667	3954.97917
6:30	4049.37	
7:00	4215.37	4132.37

Table 1: Sample of original and reduced Total Demand data for Jan 1, 2003

Time Step	QLD Total Demand	Standardized Demand
1:00	4883.145	1
2:00	4637.96833	0.949791237
3:00	4382.586665	0.897492633
4:00	4139.015	0.847612553
5:00	4029.1125	0.825106054
6:00	4016.97	0.822619439
7:00	3954.97917	0.809924581
8:00	4132.37	0.84625175
9:00	4440.82583	0.909419202
10:00	4312.964	0.88323488

Table 2: Sample of Total Demand and standardized data for Jan 1, 2003

Next, as recommended by Dywer [**1**], the Total Demand was standardized to ensure that meaningful comparisons could be conducted across regions. So, for each region the Total Demand at 1 a.m. on each day was taken as a reference value and equated to 1. Each subsequent timestep for that day was calculated as a ratio of the reference value. Table 2 provides a small sample of standardized hourly data. Such a modification was justified since we wished to visualise the rate of change from one timestep to the next across regions, not the difference in raw Total Demand data. Note that we may have chosen any given timestep as the reference value, provided we choose the same timestep for all regions. This standardization process



Snowy Data Remaining Regions Data

Figure 1: A comparison of the Total Demand data for the Snowy against the other four regions, March 2003

was not applied to the RRP data, as we wished to consider the difference in price per megawatt hour between regions.

The visualisation displayed in Figure 1 provides a snapshot of the data for Total Demand for the month of March. It clearly indicates that Total Demand for Queensland, NSW, Victoria and South Australia have a high degree of similarity, while the Snowy region's data does not follow these trends. This relationship is typical for all months of 2003 and typical for the RRP data. Therfore, in order to obtain an accurate estimate of trends in Queensland relative to other regions, we have excluded the Snowy data.

To investigate the behaviour of a single region versus the other three regions (in this case Qld vs NSW, Vic and SA) the average of the Total Demand (RRP) for the three regions NSW, Victoria and South Australia was taken for each timestep, then this value was subtracted from the corresponding Queensland value. This provided us with an estimate of how Queensland differed from the average of the selected regions. The Total Demand and RRP data was stratified into nine bands as set out in Table 3. The bands were selected to distribute the data into discrete (and visually distinct) classes. The upper and lower classes partition off the very high and very low values in the data, and the intermediate classes are intended to provided a reasonable distribution of the remaining data. However, iIbe noted there was much more variation in the RRP values, and so checks were carried out to ensure the banded data reflected the raw data.

The data was stored in either $(s_i \times 24)$ arrays, m_i for $i = 1, \ldots, 12$, and $s_i \in \{28, 29, 30,$
$31\}$, corresponding to the twelve months of the year, or in a 365×24 array containing the data for an entire year (366×24 in a leap year).

4 Visualisation

The visualisation process began with the construction a 3-D surface plot for Total Demand and RRP for each of the twelve arrays, m_i. Here the horizontal axes plot the relationship between the hour of the day and the day of the month. The difference between Queensland's Total Demand (RRP) and the average of the three selected regions was sorted into nine bands from -4 to 4 and was mapped against the vertical axis. Thus the entry z in cell (x, y) represented the value in cell (x, y) of the matrix m_i. The visualisation displayed in Figure 2, gives the corresponding surface plots for January and July, 2003. Coloured surface plots for each of the twelves months can be viewed at the website [3].

The Total Demand plots tend to indicate the presence of two separate trends. First, we can see that for all days y there is a strong tendency for the value in cells (x, y), for $x = 0, \ldots 5$, to be within the first band. Thus for the first five hours of each day Queensland's Total Demand is close to the average Total Demand for the three selected regions. However this trend does not appear to occur in the RRP data. Secondly, for both sets of data there appears to be troughs in the surface plot which may indicate days when Queensland's Total demand and RRP is uniformly close to that of the selected states. However, this second trend is not as pronounced as the first and needs further investigation. Horizontal rotation of the plots may improve the identification of these trends, but this is not possible in the current printed form. So, while this type of plot is visually interesting, much fine detail may be hidden. To investigate both of the trends, two further visualisation techniques were used.

The next technique applied to the data was the development of a discrete contour plot for each month. We defined a function $C : \{-4, \ldots, 4\} \to \mathcal{C}$, where $\mathcal{C}$ is a set of nine distinct colours. Then, for $i = 1, \ldots, 12$ and $s_i \in \{28, 29, 30, 31\}$, the data in the $s_i \times 24$ array m_i, was projected onto an $s_i \times 24$ grid, where cell (x, y) is coloured with $C(z)$ if z occupies cell (x, y) of m_i.

Figure 3 shows this plot for the months of January and July 2003, for both Total Demand and RRP data. The homogeneous nature of the first five hours of the Total

Total Demand Data x_1	RRP Data x_2	Band
$x_1 > 0.2$	$x_2 > 100$	4
$0.2 \geq x_1 > 0.15$	$100 \geq x_2 > 4$	3
$0.15 \geq x_1 > 0.1$	$4 \geq x_2 > 2$	2
$0.1 \geq x_1 > 0.05$	$2 \geq x_2 > 1$	1
$-0.05 \leq x_1 \leq 0.05$	$-1 \leq x_2 \leq 1$	0
$-0.1 \leq x_1 < -0.05$	$-2 \leq x_2 < -1$	-1
$-0.15 \leq x_1 < -0.1$	$-4 \leq x_2 < -2$	-2
$-0.2 \leq x_1 < -0.15$	$-100 \leq x_2 < -4$	-3
$x_1 < -0.2$	$x_2 < -100$	-4

Table 3: Band table Total Demand and RRP data

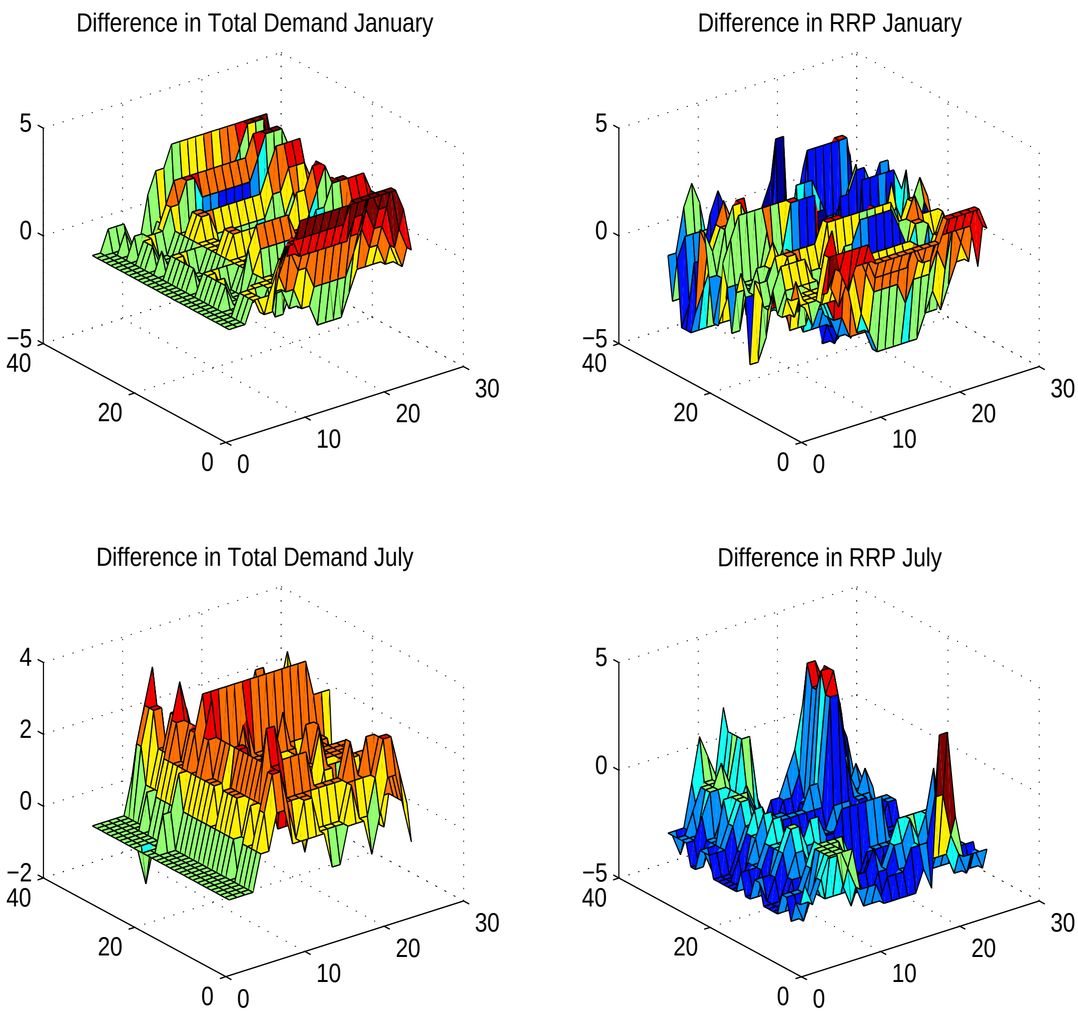


Figure 2: The Difference in Total Demand and RRP for Queensland against the other 3 regions for January and July 2003

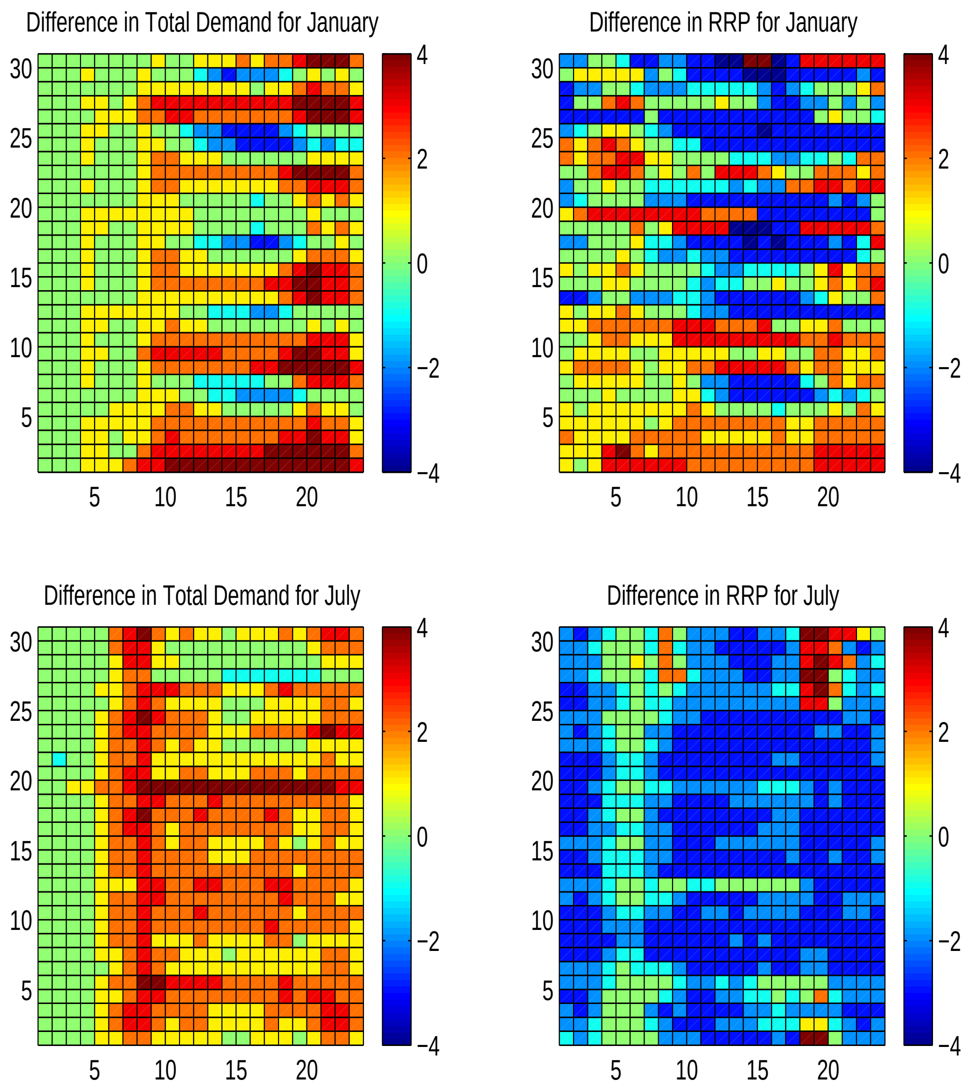


Figure 3: The Difference in Total Demand and RRP data for Queensland against the other 3 regions for January and July 2003

Demand data is clearly visible, as are the troughs. Similarly colour plots for each month of 2003 may be viewed at [3]. These plots all indicate that, from midnight to five a.m. each day, Queensland's Total Demand was close to the average of the other selected states. However, the plots also indicate that for the summer months this trend stretches from midnight to eight or nine a.m., which may correlate with an increase in daylight hours in summer. But it should also be noted that this trend is not exhibited in the evening. Interestingly, the RRP plot for January is unlike its Total Demand counterpart. The Total Demand and RRP plots for July show a little more similarity, but no obvious direct correlation between the two is apparent. As for trends across days of the week, while the contour plots show troughs they do not show regular trends corresponding to particular days of the week.

Another interesting fact arising from this visualisation, particularly for July, is that for most of the time Queensland's RRP is below the average of the three regions. As we get to the latter stages of 2003 this trend is more pronounced.

The final visualisation technique employed here is a cumulative area function, which was plotted for the monthly matrices. Figure 4 displays this plot for Total Demand and RRP data for the months of January and July 2003. Similar plots for each of the months may be view at [3].

In the cumulative area plot the x-axis represents hours and the value for any hour i may be defined as

$$\sum_{\text{each day of the month}} \text{Difference in Total Demand (RRP) data for hour } i$$

This visualisation technique again highlights the homogeneous nature of the Total Demand data for the early part of each day. As these visualisations are view across the twelve months we see that the RRP for Queensland does tend to be closer to the average RRP for the three states for hours 3 a.m. to 8 a.m. Such a trend was not as pronounced in the other two visualisations. Also the tendency of Queensland RRP data to fall below the average of the three region is far more visible with this type of plot. Finally, the cummulative area function viewed across all months of the year suggests that there may be some correlation between Total Demand value at about 7:00 a.m. and 4:00 p.m., and the RRP value at these times. The plots for Total Demand and RRP regularly show peaks and/or troughs around these times. For instance, July's Total Demand data has a spike at about 7:30 a.m. and the July's RRP data spikes just prior to this time.

5 Conclusion

Each of the above visualisation technique revealed different information about trends in the data. The surface plots suggested a regularly occurring period in each day for which the Total Demand data was close to the average of the three regions. There was a hint of troughs occurring in the surface plot corresponding to complete days of little difference in the data. The contour plots supported this supposition,

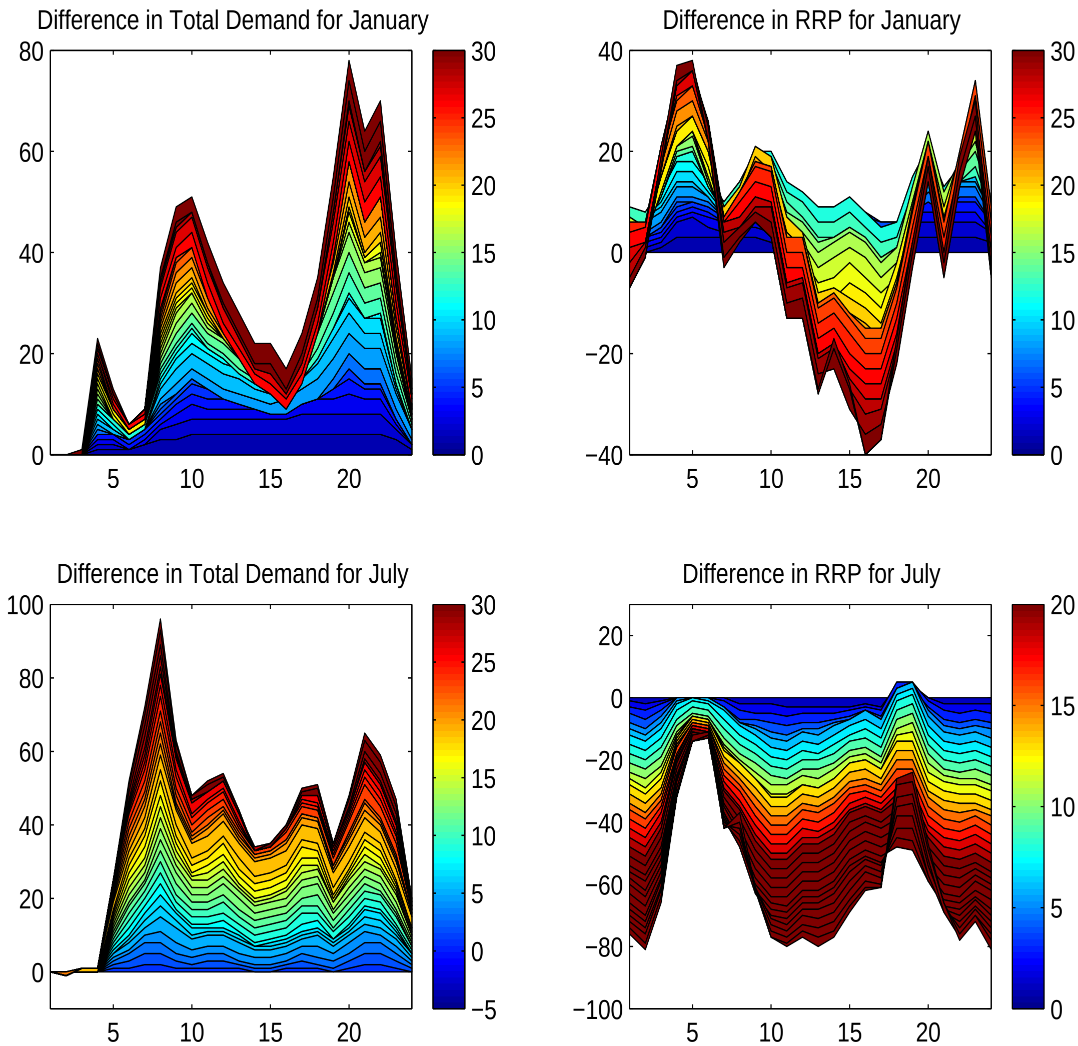


Figure 4: A cumulative area plot for Total Demand and RRP data for Queensland against the other 3 regions for January and July 2003

suggesting that the trend is affected by the month of the year and further suggested a slight correlation between Total Demand data and RRP data. However, the contour plot did not provide evidence of a regular trends on a weekly basis. Finally, the cumulative area plots suggested a better correlation between these two types of data.

This research has suggested many new directions and it is expected that future work on this problem will include a more rigorous statistical analysis of the data.

In conclusion it can be said that each visualisation technique has something interesting to add to the data analysis. Data plots of each type for all months of 2003 may be viewed at [3].

Acknowledgement The authors wish to acknowledge the assistance of Graeme Chandler and Pamela Burrage in regards to the matlab visualisations.

References

[1] Dwyer, T., and Eades, P. *Visualising a fund manager flow graph with columns and worms*. Proceedings of the Sixth International Conference on Information Visualisation (IV'02), (2002).

[2] http://www.nemmco.com.au/data/marketdata.htm (2004).

[3] http://www.maths.uq.edu.au/ dmd/energy/energydata.htm (2004).