Spreadsheets in Education (eJSiE)

Volume 2 | Issue 2 Article 2

May 2007

Excel and the Goldbach Comet

John Baker Natural Maths, john@naturalmaths.com.au

Follow this and additional works at: http://epublications.bond.edu.au/ejsie



This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

Baker, John (2007) Excel and the Goldbach Comet, *Spreadsheets in Education (eJSiE)*: Vol. 2: Iss. 2, Article 2. Available at: http://epublications.bond.edu.au/ejsie/vol2/iss2/2

This Regular Article is brought to you by the Bond Business School at ePublications@bond. It has been accepted for inclusion in Spreadsheets in Education (eJSiE) by an authorized administrator of ePublications@bond. For more information, please contact Bond University's Repository Coordinator.

Excel and the Goldbach Comet

Abstract

This article shows ways in which Excel can be used to explore the Goldbach Conjecture. We first show an elementary approach in which the Goldbach Comet is constructed by means of a table of values. Second, we give two simple VBA programs that generate substantially more data for the Goldbach Comet. We then show how the Comet can be seen as the agglomeration of individual bands and use that approach to suggest an improvemen to the Hardy-Littlewood formula for the number of ways in which an even number can be expressed as the sum of two primes. We also outline two further investigations into the form of the Goldbach Comet. After expanding the comet to show exactly which prime pairs make a given even number, conditional formatting is used to elucidate patterns in those pairings, and we also explore the possibility that Goldbach's conjecture might still hold even when the number of primes used is greatly reduced. Throughout, the focus is on ways in which Excel can be used to process and display very large amounts of data and on the underlying patters that emerge when it is used in this way.

Keywords

Spreadsheet, Visual Basic for Applications, Goldbach's Conjecture, Goldbach's Comet, Number Theory

Distribution License



This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Excel and the Goldbach Comet

John Baker Director, Natural Maths john@naturalmaths.com.au

Abstract

This article shows ways in which Excel can be used to explore the Goldbach Conjecture. We first show an elementary approach in which the Goldbach Comet is constructed by means of a table of values. Second, we give two simple VBA programs that generate substantially more data for the Goldbach Comet. We then show how the Comet can be seen as the agglomeration of individual bands and use that approach to suggest an improvement to the Hardy-Littlewood formula for the number of ways in which an even number can be expressed as the sum of two primes. We also outline two further investigations into the form of the Goldbach Comet. After expanding the comet to show exactly which prime pairs make a given even number, conditional formatting is used to elucidate patterns in those pairings, and we also explore the possibility that Goldbach's conjecture might still hold even when the number of primes used is greatly reduced.

Throughout, the focus is on ways in which Excel can be used to process and display very large amounts of data and on the underlying patterns that emerge when it is used in this way.

Keywords: Spreadsheet, Visual Basic for Applications, Goldbach's Conjecture, Goldbach's Comet, Number Theory

Introduction

How can Excel be used to explore one of the remaining dilemmas of mathematics – that Goldbach's conjecture true? There are different forms of the conjecture, but the most well-known is that proposed by Euler in 1742:

Every even number greater than 4 can be expressed as the sum of two prime numbers.

It is not a problem that is likely to yield to the kind of mathematics that is available in the classroom, but that does not mean that an exploration of the conjecture need not be undertaken at an elementary level. In particular, all of the investigations proposed in this paper would be appropriate at the Year 11-12 levels of Secondary school for students who had chosen mathematics as a preferred branch of study. Excel provides way of lifting such an investigation off the ground, by allowing students to explore patterns within large amounts of data, to write simple the programs that are needed to generate such patterns and most importantly, to create an interactive environment in which conjectures about data can be explored and tested. In Baker and Sugden [1], it was noted that:

Since many sequences are most naturally defined recursively (arithmetic, geometric, Fibonacci etc.), the spreadsheet offers a very

eJSiE 2(2):177-xxxxx

rich environment for investigation of sequences, in most instances, with no coding required. Further discussion may be found in Sugden (see [13] in this paper's references).

The use of (non-spreadsheet) technology in exploring the Goldbach conjecture is discussed by Cuoco and Levasseur [4] and Neuwirth and Arganbright [11] demonstrate the appropriateness of using spreadsheets in number theory. Our intention here is to build on earlier work and show that the spreadsheet is an appropriate tool for exploring a classical number-theoretical problem, as it enables us to test conjectures and show patterns in a way that remains accessible to the student. We will do this by:

- following the steps needed to acquire data,
- adding calculations based on that data,
- > making graphical representations of the data to elucidate patterns and
- > making the spreadsheets interactive.

In doing so, we come up against restrictions inherent to Excel and VBA, but suitable workarounds can be found, and these in themselves are informative. The power of the spreadsheet to generate and display large data sets in this way can be extended to many other situations, such as the Barnsley's [3] *Chaos Game*, as described in Baker, Hvorecky and Sugden [2] or to an investigation of what values of a, b and c will satisfy $a^2 + b^2 = c^2$ with constraints such as those suggested by Kraitchik [10].

But first, what is the Goldbach Comet? First mentioned by Fliegel and Robertson [5], the comet shows the number of ways in which an even number can be expressed as the sum of two primes. Because this number can be counted for every even number, we can define a 'comet' function:

c(n) = the number of ways that n can be expressed as the sum of two primes, where n = 6, 8, ..., an even number greater than 4.

For example,

$$6 = 3 + 3$$
, giving $c(6) = 1$
 $8 = 5 + 3$, giving $c(8) = 1$
 $22 = 3 + 19 = 5 + 17 = 11 + 11$, giving $c(22) = 3$

By systematically counting the number of ways, a graph of c(n) can be generated (see Figure 2 or the website of Herkommer [5])).

It is not hard to understand why it might be called the Goldbach Comet. The image of Halley's comet from the NASA website, Figure 1, has very similar characteristics!



Figure 1: Halley's Comet

The graph of Figure 2 raises questions such as what causes the concentrations of counts or *bands* or why might there be gaps in the distribution. For example, one

might say that there are five identifiable bands within the comet, and yet as one looks closer, there appear to be more ... and more. In contrast to the bands of concentration, there are voids. For example, there are very few even numbers between 90,000 and 100,000 that can be made in 1000 ways or thereabouts.

Figure 2 is derived from the data given in *Comet Spreadsheet 1.xls*, where the simple programs described below are used to make a list of primes from 2 to 100,000 and to calculate the comet function, c(n).

A Preliminary Investigation

To get a feel for the way in which the comet arises, students could be encouraged to create the data for c(n) on a spreadsheet that caters only for primes to 1,500. The process would be to make a table of values that has the primes 3, 5, 7, 11, ... running across the top named *prime1* and down the side named *prime2* and to use a formula such as the following to create a table of even numbers, each being the sum of two primes. The formula is:

```
=IF(prime1 < prime2, "", prime1 + prime2)</pre>
```

The limitation of having available only 256 rows only makes it possible to explore primes less than about 1,500 and the spreadsheet, *Comet Spreadsheet 0.xls*, shows the method at work. To create a table of values for c(n) for $n \le 1500$ we have named the table of even numbers as *pairs* and used the COUNTIF function to find c(n):

```
=COUNTIF(pairs, evens)
```

You will note that within *pairs*, there are even numbers greater than 1,500, but these will be disregarded by the COUNTIF function as the values for c(n) go no further than n = 1500. The graph of c(n) begins to have the characteristics of the Goldbach Comet, as Figure 3 shows. But the limitation of the spreadsheet to 256 columns means that a different approach is needed if the investigation is to be extended.

Extending the Investigation

In order to extend the investigation, we need a list of prime numbers, and preferably plenty of them. We started by visiting the Primes website [15][14], supported by the University of Tennessee at Martin, where there are lists of primes of varying length that can be downloaded. Inspection of the first 10,000 primes shows that the 10,000th prime is 104,729. This allows us to search for combinations of primes that make all even numbers up to 100,000 and since we are looking only at the even numbers, we need 50,000 storage positions, which is comfortably within the 65,536 rows available on most spreadsheets.

An alternative is to write a short program to generate primes along the lines of the Sieve of Eratosthenes; an animated example of the sieve is given by the Wiki Encyclopedia. For this, a Boolean array, varPrimes(varI), is initialised to TRUE, to indicate that we assume varI to be prime. For all multiples of varI, varPrimes(varI) is set to FALSE to show that they are composite. The following implementation of this approach is quite accessible to students:

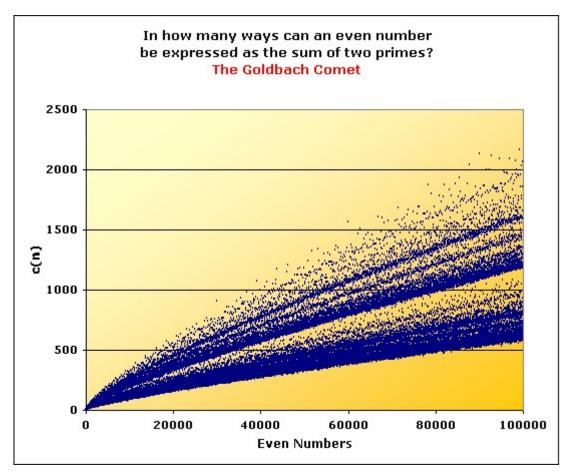


Figure 2: The Goldbach Comet

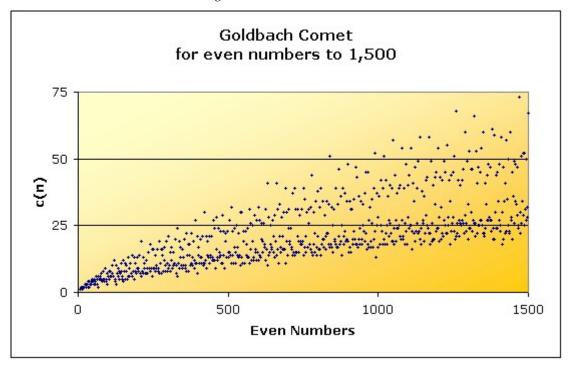


Figure 3: Goldbach Comet for even numbers to 1,500

The next step is to write a program to calculate the value of c(n) for every even number up to 100,000. We begin by defining a limit to the number of primes examined – in this case varCometLimit is 9,592, because the 9,593rd prime is the first prime greater than 100,000. We then assume that the first prime in p+q is the smaller of the two and calculate c(n), called varComet in the program, by incrementing varComet(varP + varQ) by 1 within a double loop. The details are:

```
For varI = 1 To varCometLimit
  varP = varPrimes(varI)
  For varJ = varI To varCometLimit
     varQ = varPrimes(varJ)
     If varP + varQ <= 100000 Then
          varComet(varP + varQ) = varComet(varP + varQ) + 1
     Else
          GoTo nextI 'varP + varQ is out of range
     End If
Next varJ
nextI: Next varI</pre>
```

The program creates a table of data on the *Comet List* worksheet of *Comet Spreadsheet 1.xls* in which the number of prime pairs that form each number from 2 to 100,000 has been calculated. The next step is to make a chart of these data. This is somewhat complicated by the fact that you can only chart a series of 32,000 data items, and as a result the Comet Chart has to be broken into two series.

Exploring the Comet

A useful feature of Excel is the way in which, as you hover over a data point, a tool-tip appears showing what that data point is. The diagrams in Figure 4 give two examples that come from the top end of the even numbers. The point (90090, 2135) shows that the even number 90090 can be expressed as the sum of two primes in 2,135 different ways. This is substantially more than the number 95402, which can only be expressed as the sum of two primes in 552 different ways.

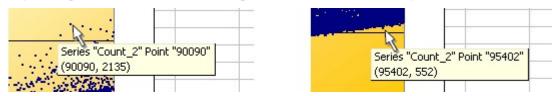


Figure 4: Pointing at part of the Comet graph

But 90090 and 95402 also differ in another striking way:

```
90090 = 2 \times 3^2 \times 5 \times 7 \times 11 \times 13

95402 = 2. \times 47701
```

That is, 90090 has 7 prime factors while 95402 has only 2. Hovering over different numbers and finding their factorisation confirmed that this type of result was not just a one-off phenomenon. The Hubbard book of factorisations [9] made it easy to locate these factors, but there is a factorising applet that is quick to use available from a cryptographic website of Hodges [8]. You can keep the applet open while you find the factors of numbers in different parts of the comet.

After exploring the factorisation of quite a few even numbers, it began to appear that the number of prime factors of an even number has a definite bearing on its comet number. Indeed the more prime factors a number contains, the larger the count of prime pairs that add to it – a statement that seems to have a sense of paradox to it.

Another feature of the exploration was not being able to find a multiple of 3 in the lowest band of the comet. On the other hand, every number tested in the second band appeared to have a factor of 3. This feature suggested the first investigation.

Bands for Even Numbers with a Common Factor

To explore how the Goldbach Comet is affected by restricting attention only to numbers that are a multiple of a chosen number, a new spreadsheet is needed (*Comet Spreadsheet 2.xls*). It can be based on the Count data calculated by the program, but it is advisable to work with another spreadsheet in order to keep the number of refresh calculations to a reasonable level. The Figure 5 shows the bands for numbers that have 3 as a common factor and it confirms the suspicion mentioned above that even multiples of 3 do not have a comet number in the lowest band.

The band for multiples of 5 is possibly even more interesting as it contains two very distinct parts, see Figure 6. The lower band has almost no overlap with the bands for multiples of 3, while the upper band appears to be fully contained within the third band of multiples of 3. This hypothesis can be tested by looking at the bands for 15, and further at the bands for 105, which both occupy the top part of the bands for 5. In their investigation of the Goldbach Comet as a source of fractals, Wang, Huang & Dai [16] also noted how the comet divides into bands associated with different multipliers.

Since the Goldbach Conjecture applies to even numbers, or multiples of 2, it is interesting to consider multiples of 2^k , that is for multiples that are powers of 2. In contrast to the band for multiples of 3 and 5, the bands for multiples of powers of 2 appear to be less dense versions of the complete comet. That is, multiples that are powers of 2 seem to have very little influence on the distribution of counts, certainly nothing as striking as the effect of viewing only multiples of 3.

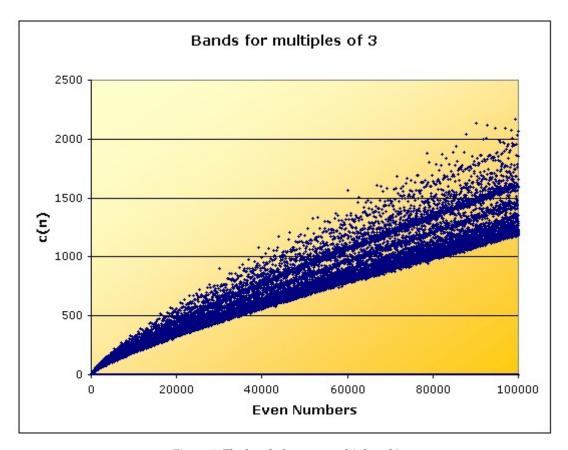


Figure 5: The bands for even multiples of 3

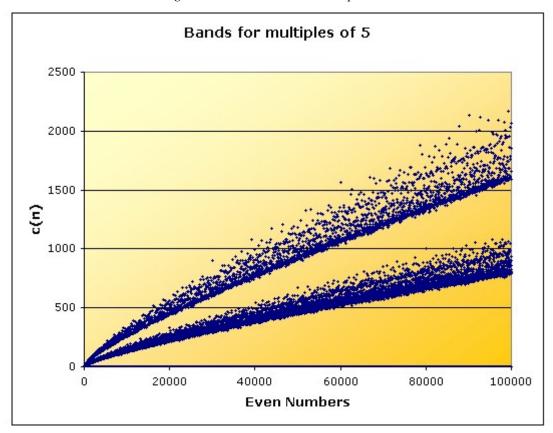


Figure 6: The differentiated bands for even multiples of 5

Prime Multiples of a Given Number

The Goldbach Comet is generated by looking at all even numbers, regardless of the number of factors that they might have. One way to control the number of factors is to look only at prime multiples of a given base number. For example, if the base number is $12 = 2^2 \times 3$, then the sequence 12×2 , 12×3 , 12×5 , 12×7 ... in which 12 is successively multiplied by a single prime number, is a sequence of numbers all of which have 4 prime factors. This time, the result is quite striking, because, for any given even base number n, the graph of $\{c(np): p \text{ is any } odd \text{ prime}\}$ forms a single band, as in Figure 7, which is derived in *Comet Spreadsheet 3.xls*.

As you will see, the range of even numbers in Figure 7 has been extended to 200,000. This was done to confirm the trends that working with numbers to 100,000 suggested. As we look down the data that generated this chart, it is clear that the values of c(n) are not exactly monotonically increasing with n. That is why the image is that of a band, rather than of a line. But the general trend is definitely for a single band that continues to separate from the horizontal axis.

The Hardy Littlewood approximation

In their paper of 1922 on the expression of a number as a sum of three primes, Hardy and Littlewood [6]¹ give an approximation² to the comet number of an even number in the form of a conjecture:

Conjecture A. Every large even number is the sum of two odd primes. The asymptotic formula for the number of representatives is

$$N_2(n) \approx 2C_2 \frac{n}{(\log n)^2} \prod_p \left(\frac{p-1}{p-2}\right)$$

where p is an odd prime divisor of n, and

$$C_2 = \prod_{\omega=3}^{\infty} \left(1 - \frac{1}{(\omega - 1)^2} \right)$$
 taken over all odd primes.

The use of the suffix 2 distinguishes the above formulae from an earlier formula giving n as the sum of three primes and C_2 is defined as a product taken over all odd primes, ω . These formulae lend themselves to evaluation on a spreadsheet, as the formula for C_2 can be simplified and written in recursive form as:

$$C_{2,n} = C_{2,n-1} \left(1 - \frac{1}{(\omega_n - 1)^2} \right)$$

¹ Although not readily accessible, this landmark paper is reproduced in Wang [17] that charts the progress to date towards a solution of the Goldbach Conjecture.

² A slightly different for of the Hardy-Littlewood approximation in which (log n)² is replaced by log(n) log(n-2) is given by Oliveira e Silva [12], whose has been a participant in the search for large numbers (up to 10¹⁷) that might not be the sum of two primes. The search continues!

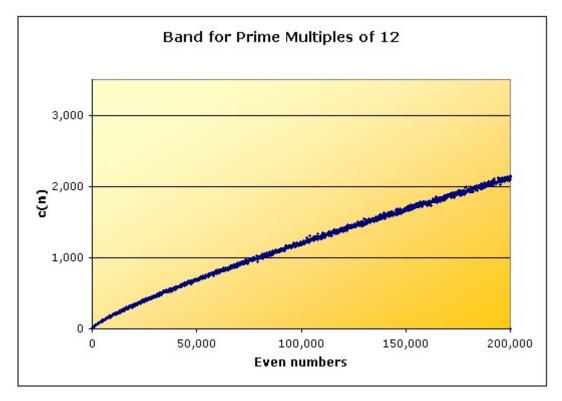


Figure 7: The band of counts for all prime multiples of 12

Using this formulation, we find that $C_2 = 0.660162$ (approximately) and the form of the product suggests that the iterative process is convergent which was proven to be the case by Wrench [18]. C_2 is also known as the *twin primes constant* and is discussed at the University of Tennessee at Martin site [14].

When we convert the formula that approximates $N_2(n)$ to spreadsheet form, the comet for prime multiples of a given base number lends itself to a similar recursive form, since the odd prime factors of the base number can be evaluated once, as:

$$P_{base} = \prod_{p} \left(\frac{p-1}{p-2} \right)$$
 where p is an odd prime divisor of the base number.

Because the prime multiples of n will be only those that divide the base number, P_{base} , and the current prime multiplier, p_n , of the base number the recursive formula for the approximation is given by:

$$N_2(n) = 2C_2 \frac{n}{(\log n)^2} P_{base} \left(\frac{p_n - 1}{p_n - 2} \right)$$

When this recursive formula is used, the Hardy-Littlewood approximation of Conjecture A can be compared with the actual comet for a given base on a graph, as in Figure 8. The spreadsheet *Comet Spreadsheet 3.xls* shows how this can be done.

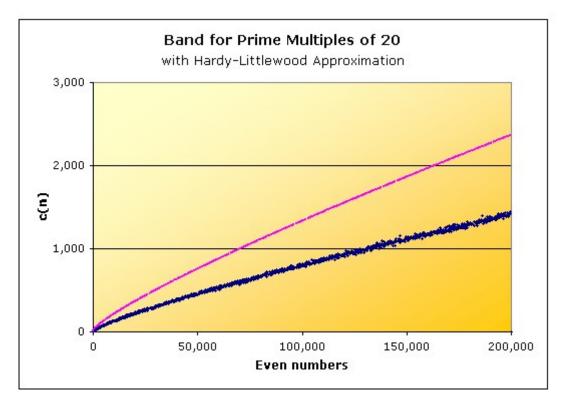


Figure 8: Comparing the Hardy-Littlewood formula with the band for prime multiples of 20

It appears that the Hardy-Littlewood approximation over-estimates the value of c(n), and this was found to be the case for every base number tested. However, when P_{base} is reduced by a factor of 3/5, the agreement between the actual and approximation is remarkably good ... at least for even numbers up to 200,000.

This reduction factor was tested for many different base numbers and appears to be correct. But why the Hardy-Littlewood formula over-estimates and why a simple reduction by a factor of 3/5 corrects the over-estimation are not so clear.

The process by which the comet for prime multiples of a given base number is created depends completely on the value assigned to P_{base} . In particular, the calculation of P_{base} needs to account for repeated primes factors. For example, the base number for $54 = 2 \times 3^3$ is calculated as follows:

$$P_{base} = \frac{3-1}{3-2} \times \frac{3}{5} = 1.2$$

That is, repeated prime factors of the base number are only used once in the calculation of P_{base} . The graph of Figure 10 shows the comet for 54 with 1.2 used as P_{base} .

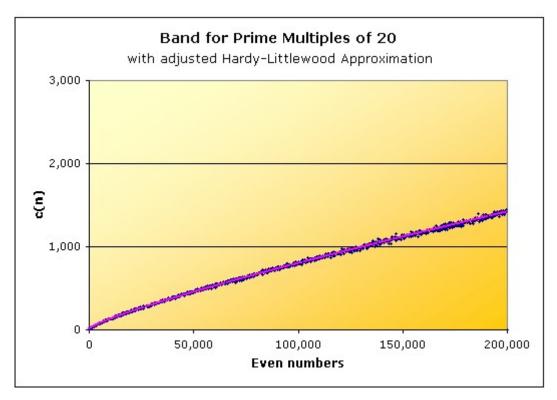


Figure 9: Comparing a revised Hardy-Littlewood formula with the band for prime multiples of 20

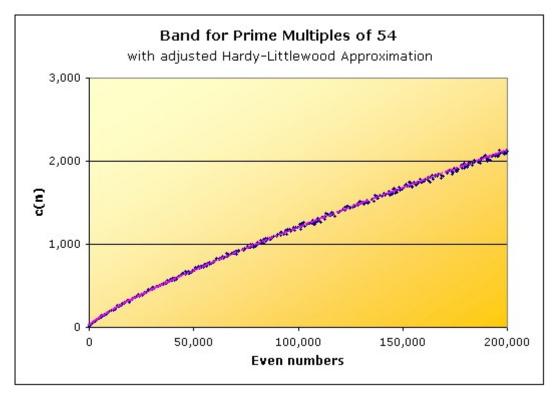


Figure 10: Comparing the revised Hardy-Littlewood formula with the band for prime multiples of 54

11

An alternative view of the Comet

The Comet function, c(n), counts the number of ways in which the even number n can be expressed as the sum of two primes. If we zoom in a bit, we can look at exactly which two primes can be used to form n. The result of doing this is shown in Figure 11, where the primes are listed across the top of the table, the evens down the side, and the body of the table shows which primes combine with those of the top row to give n. For example, the 5^{th} row of the table is for the even number 10, and it shows that 10 = 7 + 3 = 5 + 5 = 3 + 7.

The formula used in the body of the table uses the names *primes* and *evens*, which name the sides of the table, as well as the logical function ISNA, meaning 'is not available'. The formula is:

```
=IF(evens > primes, IF(ISNA(HLOOKUP(evens - primes, primes, 1, FALSE)), 0, evens -primes), 1)
```

The formula thus returns a 1 in the part of the worksheet where the even number is less than the prime in that column, otherwise it looks up the difference between the even number and the prime within the list of primes. If that difference is also prime (i.e. is found in the list), the difference is displayed, otherwise a 0 is displayed. Conditional formatting is then used to differentiate between primes (purple), zeroes (white) where the difference between the prime and the even is composite and ones (cyan) where the even number is less than the prime.

The worksheet fragment of Figure 11 covers only even numbers between 2 and 50. It is a small part of *Comet Spreadsheet 4.xls*. However, when we zoom out to 10%, the smallest allowable zoom value, the result is the diagram of Figure 12. This no longer has the look of a comet, but it could be appropriately described as the **Goldbach Glacier**.

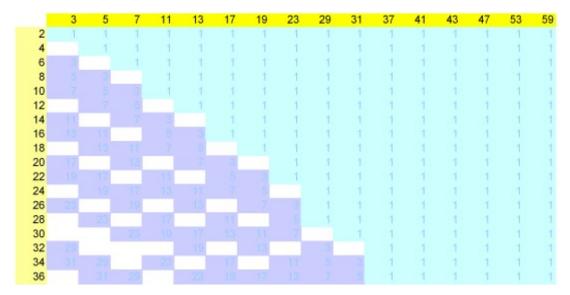


Figure 11: Listing the prime combinations that make each even number

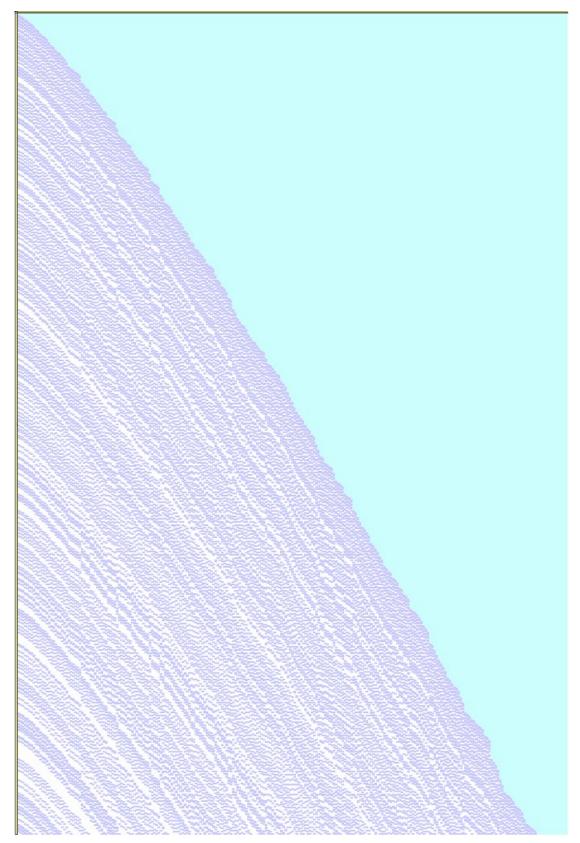


Figure 12: The Goldbach Glacier

The Goldbach Glacier has bands but these are rather different to the bands of the Comet — they appear to be caused by the gaps in the primes. For example, the white band that starts in the lower left corner of Figure 12 starts because there are no primes between 1,327 and 1,361. In contrast, the twin-prime pair 1,319 and 1,321 starts the purple band just above that white band.

Another avenue of Investigation

There are a number of striking features of the graphs that we have shown. First, it does seem that Goldbach's Conjecture is a very weak form of the real situation. To say that an even number can be expressed as the sum of two primes is an understatement, as it seems to be possible to represent large even numbers as the sum of two primes in a large number of ways, and a good approximation to just how many does seem to be given by the Hardy-Littlewood formula with a minor adjustment.

We can look to see if there is a restriction that can be placed on which primes are used to see if it is still possible to represent even numbers as the sum of two primes from the restricted set. We explored the possibility of restricting the primes to only those that form part of a twin-prime pair. That is, we include 3, 5, 7, 11, 13, 17, 19, all of which occur as one of a twin prime pair, but we excluded primes such as 23, 37 and 47 because they are not one of a prime pair. The function that counts the comet number in this restricted case is:

tp(n) = the number of ways in which an even number, n, can be expressed as the sum of two twin primes.

Figure 13 shows the distribution of counts obtained from the twin primes less than 100,000. There are 2,447 of them and, as the distribution shows, all but a handful of the even numbers to 100,000 can be expressed as the sum of two twin-primes. Those that cannot be so expressed all occur in groups of 3, such as 94, 96 and 98, which is the smallest after 0, 2 and 4, and 4,204, 4,206 and 4,208, which is the largest over all, and interestingly, the middle number of those triples is always a multiple of 3.

Unlike the Comet diagram, the distribution now resembles a swarm of bees more closely that a comet, except that there do appear to be horizontal stripes, some quite full and some quite empty of counts. To see if there was a noticeable pattern in these stripes, we created a histogram of counts, showing the number of times that a particular count occurs within the first 100,000 even numbers. The resulting histogram of Figure 14 shows two distinct distributions, which would result in there being some counts that occur quite frequently compared with others that seldom occur.

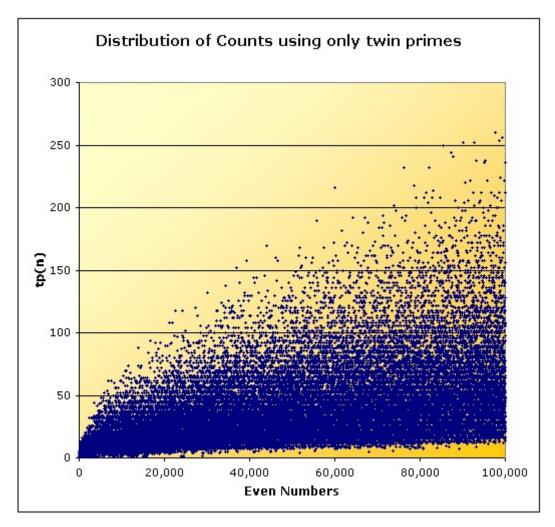


Figure 13: Distribution of counts when only twin-prime numbers are used to form the primes.

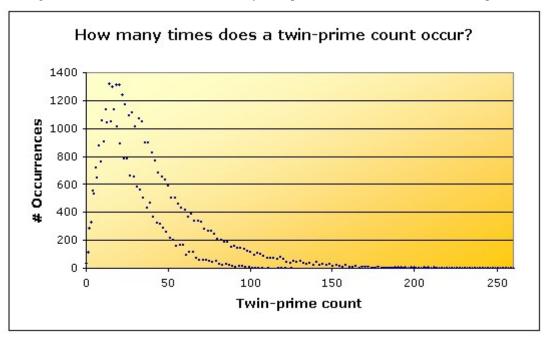


Figure 14: The distribution of counts, tp(n), that make up Figure 13

Conclusion

The aim of this paper has been to show ways in which the spreadsheet can be used to investigate one of the outstanding problems of mathematics and in a way that does not rely on an advanced understanding of mathematics. The ability of Excel to handle large datasets allowed us to find an explanation of the bands within the Goldbach Comet and to suggest a revision to the Hardy-Littlewood formula that greatly improves its accuracy. The same features allowed us to explore what happens when we restrict the primes to *twin primes*, those that differ from another prime by only 2. We also used the conditional formatting feature to see what happens when every prime pair that adds to a given even number is highlighted, with the result being the *Goldbach Glacier*. In all of these investigations, it has been the intention to demonstrate the kind of exploration of large datasets that can be undertaken with school mathematics and a good spreadsheet program.

References

- [1] Baker, J. E and Sugden, S. J (2003) *Spreadsheets in Education the First 25 Years*, eJournal for Spreadsheets in Education, Vol 1, No. 1
- [2] Baker, J. E, Hvorecky, J. and Sugden, S. J. (2005). *Recursion and Spreadsheets*, eJournal for Spreadsheets in Education, Vol 2, No. 1
- [3] Barnsley, M. F. and Rising, H (1993). *Fractals Everywhere*, 2nd ed. Boston, MA: Academic Press, 1993.
- [4] Cuoco, A., and Levasseur, K. (2003). Classical mathematics in the age of CAS. In A. Cuoco et al. (Eds), Computer Algebra Systems in Secondary School Mathematics (pp. 97-116). Reston, VA: NCTM.
- [5] Fliegel, H. F. and Robertson, D. S., (1989) *Goldbach's comet: the numbers related to Goldbach's conjecture*, J. Recreational Math., 21:1, pp 1 7
- [6] Hardy, G. H. & Littlewood, J. E. (1922) Some problems of 'Partitio Numerorum'; III: On the expression of a number as a sum of primes, Acta Mathematica, 44, pp 1 70
- [7] Herkommer, M. www.petrospec-technologies.com/Herkommer/goldbach.htm
- [8] Hodges, A. www.cryptographic.co.uk/factorise.html
- [9] Hubbard, R. L. (1975), The Factor Book, Hilton Management Services, England
- [10] Kraitchik, M. (1942), Mathematical Recreations. W. W. Norton, pp. 95-104
- [11] Neuwirth, E. & Arganbright, D (2004), *The Active Modeler: Mathematical Modeling with Microsoft Excel*, Brooks/Cole, Belmont (CA)
- [12] Oliviera e Silva, T. www.ieeta.pt/%7Etos/goldbach.html#r2
- [13] Sugden, S. J. (2002). *Illustrating Mathematical Fundamentals with Microsoft Excel*. CATE 2002, The 5th IASTED International Multi-Conference in Computers and Advanced Technology in Education, Cancun, Mexico, ACTA Press, 323—327.
- [14] University of Tennessee at Martin, primes.utm.edu/glossary/page.php/ TwinPrimeConstant.html

- [15] University of Tennessee at Martin, primes.utm.edu/lists/small
- [16] Wang L., Huang Y., & Dai Z. (2006) Fractal in the statistics of Goldbach partition, available from arvix.org.
- [17] Wang, Y. (2002) *The Goldbach Conjecture Second Edition*, World Scientific Publishing, Singapore
- [18] Wrench, Jr, J. W. (1961) Evaluation of Artin's Constant and the Twin-Prime Constant, Mathematics of Computation, Vol. 15, No. 76 (Oct., 1961), pp. 396 398