Sol Tychonis

A Case Study in Bayesian Data Analysis

Michael P. McLaughlin

December, 2021

$\Psi^*\Psi$ causaScientia

1 Introduction

Back in the day, my thesis advisor would often come into the lab, see graduate students and postdocs busily working away and declare in a loud voice, "Science never sleeps." He was right, of course, but it is not only science that never sleeps. Just about anything related to technology tends to change and improve over time—at least a little and, occasionally, quite a lot. In the latter category, albeit fairly obscure, is the methodology of data analysis.

Unless you have to do it for a living, the process of analyzing data is an activity that is so much taken for granted that it seldom if ever receives any attention. One hears about it and knows that there is a discipline called *statistics* that people use to accomplish it somehow. Apart from that, it is a black box which hardly anyone cares to open. That traditional statistics became obsolete at the end of the last century is certainly not common knowledge.

Traditional, *frequentist* statistics begins by making some assumptions about the data. It then applies some numerical test(s) to the dataset. Finally, after making one or more further assumptions, it converts the value(s) output by those test(s) to a probability, termed a *p-value*, intended to quantify the uncertainty in any information gleaned from the data. Most data analysis in the twentieth century was accomplished in this fashion even though it was well understood that it was just an elaborate workaround for the underlying problem for which the math was much too difficult to solve analytically. This is not to say that the answers obtained from statistics were invalid; on the contrary, they appeared to work very well. Still, for knowledgeable practitioners, it was hard to ignore the echo of an old saying, "When all you've got is a hammer, everything looks like a nail."

The current state of the art in data analysis does not employ statistics of any kind; instead, it utilizes *Bayesian* methodology. This is not the place to describe the Bayesian approach in detail. There are several good sources available for that. [7][9][12][10]

In general, Bayesian inference goes back to first principles, formulating everything in terms of probability from the outset. With this procedure, one develops a probability *model* for whatever was known about the problem, if anything, before seeing the data to be analyzed (*prior information*) and another for the observations themselves (*likelihood*). This allows one to capture all information relevant to the analysis and outputs results that are far more specific and useful than those afforded by traditional statistics. Moreover, it does all of this without making any assumptions or approximations apart from any that the analyst might choose to add. Also, it has the capability to quantify model comparison in a robust mathematical fashion, again without making any assumptions. The net result is a methodology that is extremely powerful, intuitive and useful.

The case study documented here cannot even begin to exhibit the full power of Bayesian data analysis. Rather, it discusses just one example demonstrating the features described above—an interesting and noteworthy problem that has been investigated previously but without the benefit of Bayesian methodology.

2 Background

For three decades of the sixteenth century, Frederick II was king of Denmark and Norway. He is not particularly famous except, perhaps, in Denmark and Norway but he should be. During his reign, he did one thing that fundamentally changed all of human civilization forever. He granted a lot of money plus the island of Hven to his compatriot and fellow aristocrat Tycho Brahe to build and operate an astronomical observatory that would be named Uraniborg.

Tycho Brahe was a man obsessed with the idea that the only way to achieve a true understanding of the world was by making the most accurate observations possible. Before Tycho, people interested in nature would often make observations but their heart was not really in it. Tradition favored theorizing over experimentation so available data were good but not great. Tycho very much wanted to be great and to be remembered for his efforts.

The data amassed by Tycho and his assistants over the years improved upon previous data to such a degree that Johannes Kepler could later use them to derive his three Laws. These Laws gave Isaac Newton something to explain in his monumental work commonly known as *Principia Mathematica*. The new understanding of physics provided by Newton led directly to the Industrial Revolution which led, in turn, to the technological civilization which we currently enjoy.

The focus on accuracy that Tycho exhibited has resulted in a number of investigations attempting to assess that accuracy. Tycho died in 1601 just as the telescope was being invented so all of his data are naked-eye observations. It is a matter of interest not only historically but scientifically to ascertain just how accurate such observations could be and how they were affected by the uncertainty in the scientific models of his day. Here, we present a new, Bayesian analysis of one large and fairly homogeneous dataset from Tycho's logs. It should be apparent that Bayesian methodology has had much the same effect on data analysis that the telescope has had on astronomy.

2.1 Problem

The problem was that, in those days, nearly every phenomenon in astronomy occurred as a function of time and nobody had a decent clock. Ultimately, the only "clock" that could be trusted was the sun which seemed to move periodically, Copernicus notwithstanding. To use the sun to tell time, however, required an accurate model for the sun's position. An accurate position could then be converted into an accurate time.

One of Tycho's routine tasks, carried out almost daily over many years, was measuring the transit altitude of the sun at Uraniborg, that is, the vertical angle above the southern horizon at local noon when the sun was highest in the sky. Since this altitude varied with a period of one year, altitude and time were related. To get other measurements as accurate as he desired, Tycho had to improve his solar model and thus his time measurements. For this, world-class solar transit altitudes would provide the necessary data provided that you knew the latitude of Uraniborg which Tycho did. But that's another story.

Our primary goal in this analysis is to assess the accuracy of Tycho's instruments and to compare them with each other, an analysis that has been done before. [13][15][14]

2.2 Instrumentation, etc.

By 1588,[1] Tycho had finished building his four main instruments the most famous of which was his great mural quadrant illustrated in Figure 1. The others were a large wooden sextant, a revolving quadrant and a large steel quadrant. Descriptions of all of Tycho's instruments are available here and elsewhere, including the original book, *Astronomiae Instauratae Mechanica*, written by Tycho himself. [3]

These instruments were quite large in order to measure angles with high resolution. They typically featured a special *transversal* scale for interpolation, a precursor to the Vernier scale invented in the following century. Also, to prevent looking through the instrument in an off-centered fashion, there were *pinnacidia*, a device intended to eliminate any parallax within the instrument. Most instruments were installed in Uraniborg but there was a smaller observatory, Stjerneborg, nearby and partly underground that was often used.

In order to get accurate results, there were also two necessary corrections to the raw altitudes (that Tycho knew about). One was refraction, an effect that made the altitude appear too high. Tycho developed several versions of his refraction table. The one shown at the right was specifically for the sun. [2, pg. 64]

There was a more subtle correction needed because models predicted the positions of body centers but all observations were made from the surface of the Earth, not its center. This induced a form of parallax which could be corrected but only if you knew the radius of the Earth (another source of error).

In his calculations, Tycho attempted to correct for refraction and parallax but there were two more very small sources of error that he did not know about:

TABVLA REFRACTIONVM SOLARIVM

Alt. ⊙ G.	Refractio ′	″	Alt. ⊙ G.	Refractio ′	″
0	34	0	23	3	10
1	26	0	24	2	50
2	20	0	25	2	30
3	17	0	26	2	15
4	15	30	27	2	0
5	14	30	28	1	45
6	13	30	29	1	35
7	12	45	30	1	25
8	11	15	31	1	15
9	10	30	32	1	5
10	10	0	33	0	55
11	9	30	34	0	45
12	9	0	35	0	35
13	8	30	36	0	30
14	8	0	37	0	25
15	7	30	38	0	20
16	7	0	39	0	15
17	6	30	40	0	10
18	5	45	41	0	9
19	5	0	42	0	8
20	4	30	43	0	7
21	4	0	44	0	6
22	3	30	45	0	5

Nutation is due to the wobble of the Earth about its axis.

Aberration is an effect due to the speed of the Earth revolving about the sun combined with the speed of light coming through the instrument.

Corrections for nutation and aberration are too small to be noticed by anyone without a really good telescope, not even an observer as expert as Tycho Brahe. Nevertheless, when we present our results in Section 5, they will be based on ground truth that takes into account all of these corrections in order to avoid adding any errors of our own.

[1] All referenced dates refer to the Julian calendar.

Figure 1: Mural Quadrant of Tycho Brahe (*Tichonicus*) [16]

3 Data

There are two sources of data: Tycho's raw observations and the true answers—what Tycho *should have observed*. We define error[i] = observed[i] − true[i] (in arc-minutes).

3.1 Tycho's Data

Tycho Brahe and his assistants kept meticulous logs of their observations along with a lot of notes and sample calculations using these data. [4] The first page of his (edited) logs for 1589 is shown in Figure 2. [5, pg. 311] (see Appendix A) We include in our dataset all solar transit altitude measurements, using the four instruments described above, for the years 1588–1595 with two exceptions:

- Observations that were recorded as dubious or otherwise uncertain.
- Fairly obvious outliers, apparently errors from assigning an incorrect date to the observation[2] (see Supplementary Material).

This is a large dataset (N = 1,432) as shown in Table 1.

Table 1: Tycho's Solar Transit Altitudes [1588–1595]

Instrument	Sample Size								
	1588	1589	1590	1591	1592	1593	1594	1595	Total
Mural Quadrant	103	138	142	23	0	3	7	18	434
Sextant	36	22	1	0	0	0	0	0	59
Revolving Quadrant	77	65	107	63	44	16	32	63	467
Steel Quadrant	63	84	107	62	43	17	32	64	472
Total	279	309	357	148	87	36	71	145	**1,432**

3.2 Ground Truth

In order to obtain correct values for Tycho's observations for the time and place where they were made, we shall use the current models for the motion of the Earth on its axis and around the sun plus a good model for refraction. [11] The first two of these are far more accurate than we require. For refraction, we cannot do as well but we can do well enough. Refraction is a function of atmospheric temperature and pressure for which we have no Uraniborg data so we must adopt average values.[3] Fortunately, the errors in doing so are very small and will not impact our conclusions.

For this study, we implemented the free AA+ library published by Naughter.

[2] most involving several instruments simultaneously

[3] 10 Celsius and 1,010 millibars, respectively

OBSERUATIONES ANNI 1589.

OBSERUATIONES SOLIS.

DIE 3 JANUARIJ[1]

Alt. $\odot$ Merid. per Chalyb.		*12*	*$42\frac{1}{3}$*
per Volub.		*12*	*41*
per Sext.		*12*	*$42\frac{1}{4}$* vel $\frac{1}{3}$
per Tychon.	*vtroque*	*12*	*$41\frac{1}{2}$*
Declin. per Armillas,	*vno*	*21*	*23*
	alt. pinn.	*21*	*$23\frac{1}{2}$*
Repetita declinatio		*21*	*$23\frac{1}{6}$*
	alt.	*21*	*$23\frac{1}{2}$*
	medium	*21*	*$23\frac{1}{3}$*

DIE 6 JANUARIJ[1].

H. $6\frac{1}{2}$ Horologium verificabatur.
H. 8 6′40″ $\odot$ ſupremus limb. oriebatur
8 9 55 Medius $\odot$ ortus.
8 11 35 $\odot$ totus ortus.

In Meridie ſequenti horologium $8\frac{1}{2}$ Min. iuſto tardius mouebatur ab Hora $6\frac{1}{2}$ matutina.

Alt. $\odot$ Merid. per Chalyb.		13	$14\frac{2}{3}$
per Volub.		13	$13\frac{1}{2}$
per Sext. nou.		13	$14\frac{2}{3}$
Declin. per	vno pin.	20	51
Armill. ſubterr.	alt.	20	$51\frac{1}{4}$

DIE 7 JANUARIJ[1].

H. 5 horologium rectificatum eſt ad Spicam ♍ & cor ♌ & aliquid ponderis ei iniectum eſt.
H. 8 5′ 20″ $\odot$ oriebatur ἐν πλάτει.
8 10 0 $\odot$ totus ſupra Horizontem eleuatus.

Eodem die in Meridie horologium mouebatur iuſto tardius $5\frac{1}{2}$ M., qui error irrepſit ab Hora 5 matutina.

Alt. $\odot$ Merid. per Volub.		13	$25\frac{2}{3}$
per Sext.		13	$26\frac{1}{2}$
per Chalyb.		13	26 25
per Mural.	vtroque	13	$25\frac{3}{4}$
Declin. per Armill.		20	$39\frac{1}{2}$
	alt.	20	$39\frac{1}{4}$

Deinde P. M. Sole verſus occaſum inclinato obſeruabatur eius declinatio in certa altitudine vt ſequitur pro refractione eruenda.

H. 2 50′	*Alt. $\odot$ 5 30*	*Decl.*	*20 $33\frac{1}{2}$*
2 55	*5 0*		*20 $32\frac{1}{2}$*
3 2	*4 16*		*20 $31\frac{2}{3}$* bona

Idque ſatis correſpondet antecedentibus ante Meridiem habitis obſeruationibus, nam mutatio eſt M. 3 quam etiam declinatio diurna permutata cauſari poterit per Horas 6.

DIE 8 JANUARIJ.

Alt. $\odot$ Merid. per Chalyb.		13	$38\frac{1}{2}$
per Volub.		13	37
per Tychon.	vtroque	13	$38\frac{1}{12}$
Decl. $\odot$ per Armill.		20	$27\frac{1}{2}$
	alt.	20	28

N. B. Horol. in Meridie correctum eſt.

DIE 9 JANUARIJ.

Alt. $\odot$ Merid. per Chalyb.		13	$50\frac{1}{3}$
per Volub.		13	$49\frac{1}{3}$
per Mural.		13	$50\frac{1}{3}$
Decl. $\odot$ per Armill.		20	15
	alt.	20	$15\frac{1}{2}$

Non erat ſatis ſerenum.

DIE 21 JANUARIJ.

Alt. $\odot$ Merid. per Chalyb.		16	49
per Sext.		16	$48\frac{5}{6}$
per Volub.		16	$48\frac{1}{2}$
per Mural.	vtroque	16	$48\frac{2}{3}$
Declin. $\odot$		17	$16\frac{5}{6}$
		17	$16\frac{3}{4}$

Alt. obſeruata	*16*	*49*	*0*	Melior
Refractio ſubtrahit		*6*	*35*	
Parallaxis addit		*2*	*53*	
Vera Alt.	*16*	*45*	*18*	
Declin.	*17*	*19*	*57*	
Long. $\odot$	*11*	*$43\frac{1}{3}$ ♒*		
Ephemerides noſtræ	*11*	*45*		

Forte non erat ſatis ſerenum, vel paulatim nimium ob refractionem ſubductum. Sequens obſeruatio certior eſt.

[1] Obſeruationes matutinæ declinationis Solis inter obſeruationes Veneris reperiuntur.

Figure 2: A Page from Tycho's Logs

4 Bayesian Analysis

There are many questions that could be asked in any analysis of these data but we shall focus on the most immediate:

1. How accurate are Tycho's solar transit altitudes using these four instruments?

2. What is the uncertainty in our answer to #1?

3. With regard to accuracy, should we distinguish these instruments or is it better (more credible) to pool them together?

4. What is the uncertainty in our answer to #3?

4.1 Procedure

In a Bayesian analysis, one begins by quantifying prior information, that is, any and all information relevant to the problem that was known before looking at some new data. This qualification is crucial. Bayes' Rule[4] makes an explicit distinction between what is *prior* (before seeing the data) and what is *posterior* (after seeing the data). Any violation of this precept invalidates the results of the analysis. In this study, we shall acknowledge that both Tycho and Kepler were convinced that these data are accurate to roughly one minute of arc[5] but allow for the possibility that they were wrong about that. We shall assume no further prior information.

The uncertainty for each component of prior information is described by its probability distribution. The likelihood of the data is described in the same way. When all of these distributions are inserted into Bayes' Rule, the mathematical formulation results in a multi-dimensional integral that is almost always intractable. In other words, it cannot be done analytically. However, an analytical form for the posterior distribution (the answer) is rarely needed provided that one can draw a large enough sample from it and, by the end of the twentieth century, the necessary algorithms plus computers powerful enough to implement them made such an approach practicable. This (simulation) technique is called Markov-chain Monte Carlo (MCMC) and is now well-known. [1][8]

MCMC simulates one or more "walkers" that move through the multi-dimensional parameter space[6] over and over, visiting different *states* (parameter vectors) repeatedly. Once equilibrium has been achieved, the walkers move through this space at a constant speed, entering and leaving states at the same rate, so that the final frequency of states visited is proportional to their inherent posterior probability. Conclusions drawn are based on the resulting, empirical posterior distribution.

We shall use MCMC to answer the four questions above.

[4] in essence, the Product Rule of probability [10, pg. 92]

[5] The data were recorded with a precision of five seconds of arc.

[6] specified by the RHS of Bayes' Rule

4.2 Models

This analysis involves a comparison of two models. The first model assumes that Tycho's four instruments were all equally accurate; the second allows for the possibility (but *not* the requirement) that their accuracies were different. We shall show how MCMC models are constructed and used. The model is typically implemented as a plain text file, with a special format, input to MCMC software (see Supplementary Material).

4.2.1 Model 1

The contents of the Model 1 file is shown below with line numbers added. The syntax here is specific to MacMCMC, the software used in this study. Other MCMC software packages will require something similar.

Model 1: Instruments Indistinguishable

```
1  Constants:
2  N = 1432, // # of points
3  qVar = 5.787e-4; // quantization variance
4  Data:
5  time[N], instrument[N], error[N];
6  Variables:
7  mean, obsVar, sigma, i;
8  Priors:
9  mean ~ Uniform(-10, 10);
10 obsVar ~ Gamma(3, 10);
11 Likelihood:
12 for (i, 1:N) {
13    error[i] ~ Normal(mean, sqrt(obsVar+qVar));
14 }
15 Extras:
16 sigma = sqrt(obsVar);
17 Monitored:
18 mean, sigma;
```

Much of this model is obvious and somewhat reminiscent of the C language.

Lines 1–3 Numerical constants including $qVar$, the variance associated with 5-arc-second precision given data in arc-minutes. This source of error will be added to the larger measurement error.[7]

[7] Variances from independent sources are additive.

Lines 4–5 Data variables input as column vectors in another text file. Instruments are encoded as integers 1–4 in the same order as described in Section 2.2. $time[N]$ is included in the data file but will not be used. Likewise, $instrument[N]$ is not used in this model but will be used in Model 2.

Lines 6–7 Non-data variables including $mean$ and $obsVar$, parameters of the subsequent likelihood expression.

Lines 8–10 Prior distributions characterizing available prior information. Since the latter is negligible, these particular priors are deliberately *vague*. (see below)

Lines 11–14 Likelihood distribution characterizing the observations in the data. With no prior information to tell us any better, this is just the Normal (Gaussian) distribution. The Central Limit Theorem and the Principle of Maximum Entropy are sufficient justification.

Lines 15–16 An optional statement computing a standard deviation from a variance.

Lines 17–18 List of parameters and any other derived quantities, to be saved in the *trace* file—one entry for every n^{th} state visited (to avoid autocorrelation).

The two priors warrant further clarification. Priors in general encode prior information and, even when we have none, the Bayesian procedure still requires us to say so. In any case, all prior information is described as a set of distributions quantifying the respective uncertainties. Priors must always be physically reasonable and, if vague (non-informative), must not enforce any constraints that the data cannot easily contradict.

Given what little we knew in advance, the Uniform prior for $mean$ is very broad, far more so than we expect to require. The prior for $obsVar$ is a Gamma distribution which describes values greater than or equal to zero. Here, it has a $shape$ parameter = 3 which means that it decreases gradually to zero. Its $scale$ parameter determines how broad it is. The mean of a Gamma distribution = $shape \times scale$, here = 30. This should be much larger than we require. When priors are meant to be non-informative, they should always be compared to the resulting *marginals* to ascertain that these priors were indeed as vague as intended. (see sect. 5)

With this elaboration, it should be apparent that the output from Model 1 will quantify *what we can say* about the accuracy of Tycho's four instruments *given* our hypothesis that, with regard to accuracy, they are indistinguishable.

4.2.2 Model 2

Model 2 is shown below. It differs from Model 1 by permitting the four instruments to be distinguished (have different accuracies). Mathematically, this is done by defining an array of $mean$ and $obsVar$ parameters (line 7), one for each instrument as indexed by the $instrument$ value in each datapoint. This does not *force* these parameters to be different.

Whether or not any of them turn out to be different will be decided by the data and prior information, as always—in this case, by the data alone since that is all that we have.

Model 2: Instruments Distinguishable

```
1  Constants:
2  N = 1432, // # of points
3  qVar = 5.787e-4; // quantization variance
4  Data:
5  time[N], instrument[N], error[N];
6  Variables:
7  mean[4], obsVar[4], sigma[4], mu0, sig0, scale, i, k;
8  Priors:
9  mu0 ~ Uniform(-10, 10);
10 sig0 ~ Gamma(3, 1);
11 scale ~ Jeffreys(0.01, 10);
12 for (k, 1:4) {
13    mean[k] ~ Normal(mu0, sig0);
14    obsVar[k] ~ Gamma(3, scale);
15 }
16 Likelihood:
17 for (i, 1:N) {
18    k = instrument[i];
19    error[i] ~ Normal(mean[k], sqrt(obsVar[k]+qVar));
20 }
21 Extras:
22 for (k, 1:4) {
23    sigma[k] = sqrt(obsVar[k]);
24 }
25 Monitored:
26 mean[], sigma[], mu0, sig0, scale;
```

Every unknown must have a prior so we require eight priors for these eight parameters. We could just make them the same as in Model 1 but that would be suboptimal; when comparing more than two unknowns, there is a better solution. [6] Here, these means and variances refer to a set of instruments which are all of comparable quality so their accuracies should exhibit some "family resemblance". We implement this by giving them a common "ancestry" encoded as *hyperparameters*: $mu0$, $sig0$ and $scale$ (line 7). The latter will influence, indirectly, the parameters that appear in the likelihood. $mu0$ and $sig0$ have priors similar to their analogues in Model 1. The Jeffreys prior (line 11) for the scale of the Gamma prior in line 14 is extremely vague. The Jeffreys distribution is Uniform in log space and is often used when even the order of magnitude of an unknown is uncertain.

The presence of hyperparameters in Model 2 makes it *hierarchical* for reasons evident in its *directed acyclic graph (DAG)*, shown in Figure 3. In practice, most Bayesian models will be hierarchical to some degree because there is usually at least some prior information available regarding how the uncertainties of the likelihood parameters should be described and the provenance of those uncertainties. Quite often, there is even information about what the hyperparameter values should be so the hierarchy could extend to several layers.

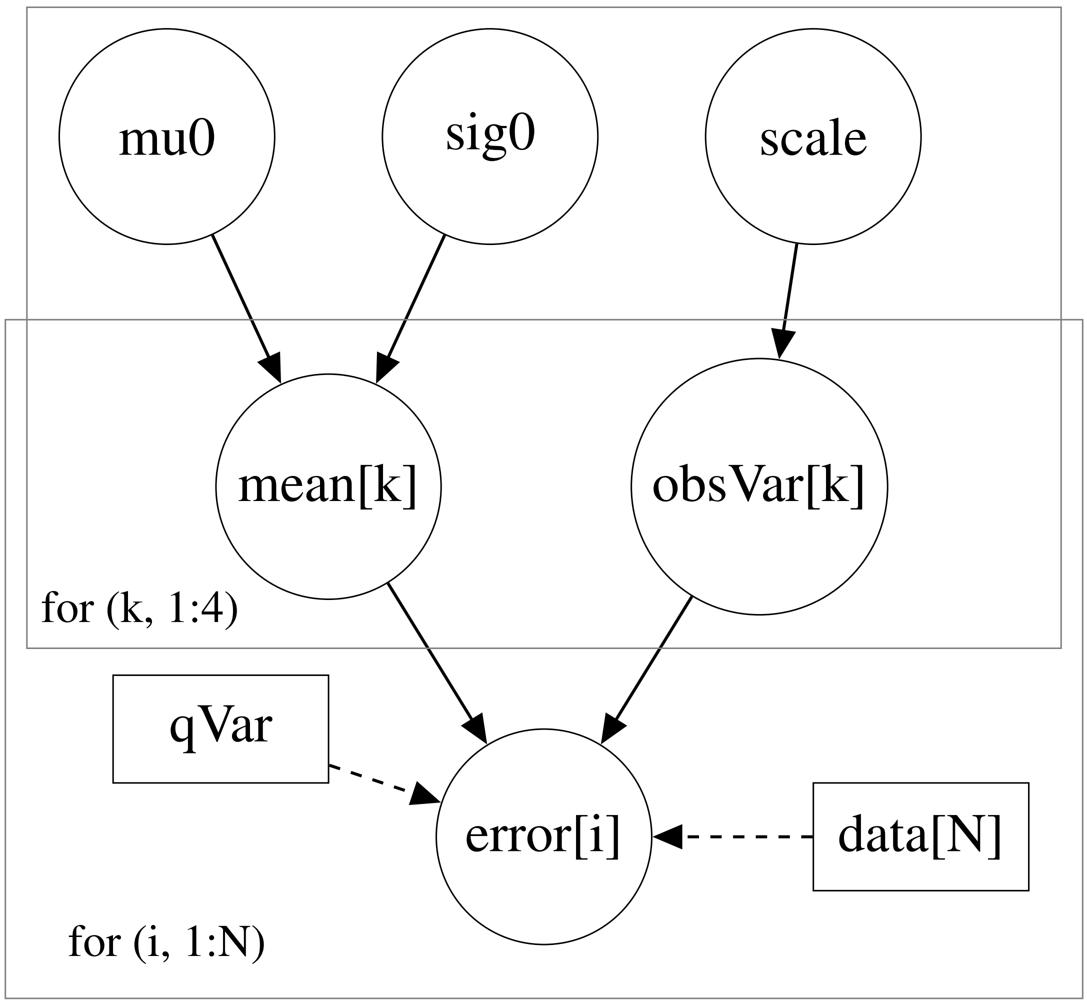

Figure 3: DAG for Model 2

As stated earlier, Bayesian analysis eschews statistics, employing probability directly. Both of our models illustrate this; they consist of a description of what is unknown, the probability distributions that quantify the uncertainties of the unknowns and how all of these are related to each other and to the data. What sort of results one obtains from a Bayesian analysis and how one makes a robust mathematical comparison of competing models will be discussed in the next Section.

5 Results

The first two questions asked in Section 4 can be answered by examining the posterior marginal for the $mean$ parameter(s) using Model 1 or Model 2, depending on the starting hypothesis. Model (hypothesis) comparison will require a computation not yet discussed.

5.1 Marginals

Each marginal is saved as a column vector in the trace file. It constitutes a (randomized) frequency histogram, in tabular form, for the corresponding unknown.

5.1.1 Model 1

In this MCMC run, there were 70 walkers.[8] The trace file contained 100,030 states (after saving every tenth state visited). Thus, the set of 100,030 values for parameter $mean$ comprise a large sample from its empirical, posterior distribution from which we can draw valid conclusions. There would have been a similar marginal for $obsVar$ but it was not saved; the derived quantity $sigma$ was saved in its place. Table 2 summarizes the marginals for the unknowns in Model 1.

Table 2: Results for Model 1

Unknown	Estimate		95% Credible Interval	
	MAP	Mean	Lower Limit	Upper Limit
mean	-0.617	-0.616	-0.648	-0.590
sigma	0.610	0.611	0.588	0.633

The answer to Question #1 is -0.616 arc-minutes, the mean of the marginal for $mean$. Therefore, the average error of Tycho's observations using these instruments was less than one minute of arc, just as he claimed. It is also negative, something that could have been predicted since Tycho was trying to measure a maximum. Even the smallest error in timing the transit would have resulted in an altitude value that was too small.

In addition to mean values, it is customary for MCMC software to save, separately, the best parameter vector found, the so-called *maximum a posteriori (MAP)* state, since it may not be in the trace file. Here, the MAP and Mean values for *mean* are nearly equal indicating that its marginal distribution is symmetric as can be seen from a plot (Fig. 4). Note that this (and all other marginals in this study) is very different from its prior.

The answer to Question #2 can be determined directly by sorting the marginal for the *mean* parameter. One can then obtain Bayesian *credible intervals* analogous to frequentist *confidence intervals* but without all of the assumptions and caveats that spoil the latter. In this case, we can say that there is a 95-percent probability that the average measurement error is between -0.648 and -0.590 minutes—very little uncertainty, in other words.

[8](5 walkers per unknown) $\times$ (2 unknowns) $\times$ (7 subprocesses)

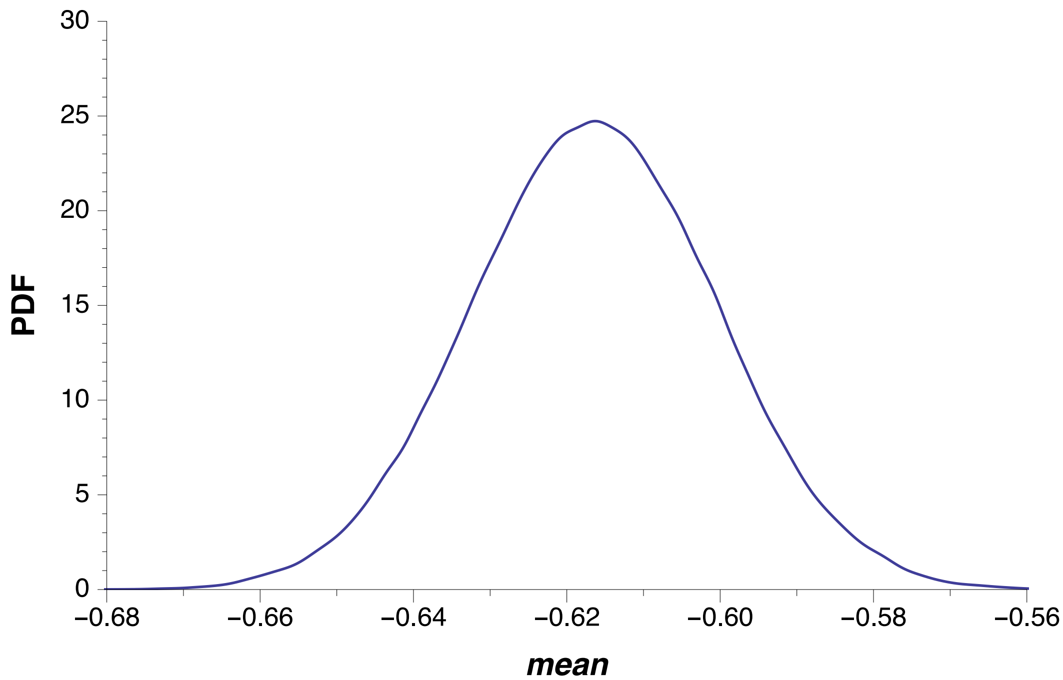

Figure 4: Model 1: Marginal for $mean$

5.1.2 Model 2

Recall that the four instruments are encoded as follows:

Mural Q. = 1, Sextant = 2, Revolving Q. = 3, Steel Q. = 4

Using Model 2 with this encoding, we can obtain a (possibly different) value for *mean* for each instrument. In fact, this model *does* seem to suggest that the four instruments are not equally accurate (see Table 3 and Figure 5).

Looking at the four curves in Figure 5 and taking them at face value, we might feel justified in asserting, for instance, that the Mural Quadrant was more accurate than all of the others even though there is a fair amount of overlap among the marginals. The accuracy of the Sextant is especially uncertain but that is understandable since its sample size is the smallest. Nevertheless, it would appear that these Model 2 results provide answers to Questions #3 and #4. As to *why* the instruments were of differing accuracy is a fair question but not one that our models could answer.

Two important issues remain. First, we have to show that these models are a good fit to the data; if not, then they are not models of the data at all. We shall address this issue in Section 5.2. Second, Model 2 gives results that *look* convincing but that does not *prove* that Model 2 is better (more credible) than Model 1. The relative credibility of the two models is something that will have to be assessed. If it turns out that the two models are equally credible (probable), then the apparent differences between the four instruments is spurious and not to be taken seriously. Model comparison will be discussed in Section 5.3.

Table 3: Results for Model 2

Unknown	Estimate		95% Credible Interval	
	MAP	Mean	Lower Limit	Upper Limit
mean[1]	−0.553	−0.546	−0.600	−0.492
mean[2]	−0.597	−0.591	−0.699	−0.480
mean[3]	−0.690	−0.704	−0.768	−0.640
mean[4]	−0.599	−0.598	−0.647	−0.549
sigma[1]	0.574	0.577	0.540	0.616
sigma[2]	0.422	0.438	0.361	0.521
sigma[3]	0.698	0.702	0.657	0.747
sigma[4]	0.543	0.546	0.512	0.581
mu0	−0.610	−0.610	−1.392	0.187
sig0	0.069	0.525	0.028	1.676
scale	0.099	0.120	0.059	0.196

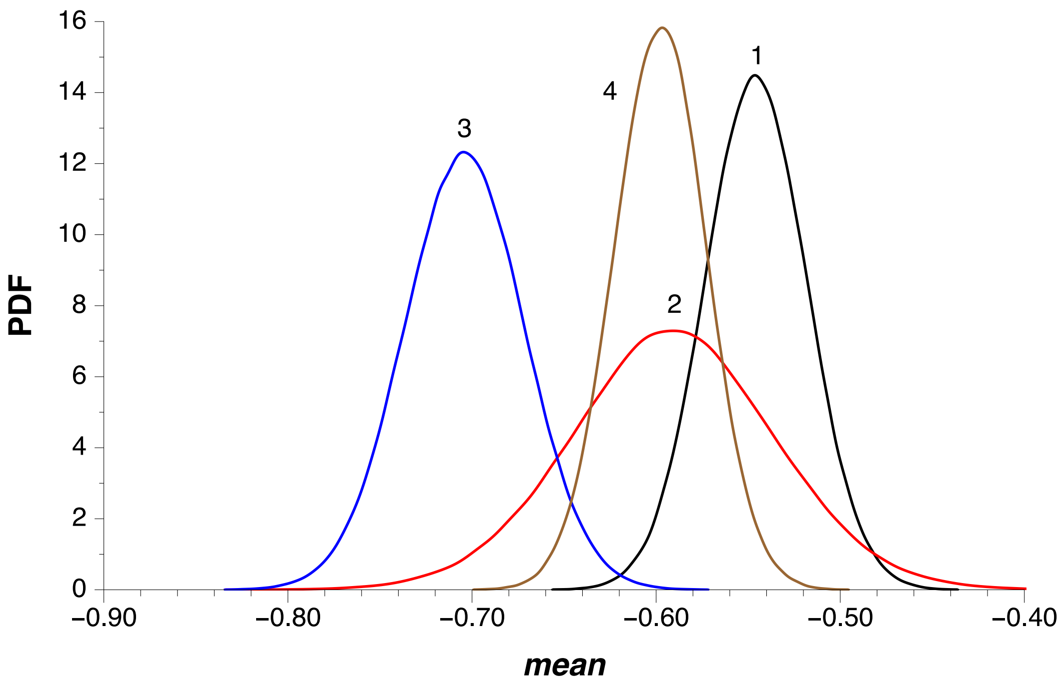

Figure 5: Model 2: Marginals for *mean*[1–4]

5.2 Goodness-of-fit

None of this analysis will be credible unless the Likelihood expression in our models is actually valid. The Central Limit Theorem says that it should be but no analysis is complete without demonstrating satisfactory *goodness-of-fit* by comparing what the data say versus what the model says. One very simple way to make this assessment is to plot the theoretical CDF (predictions) of the model versus the empirical CDF of the data.[9] If they match, we should get a straight line with slope = 1. Using Model 1, this plot is shown in Figure 6.

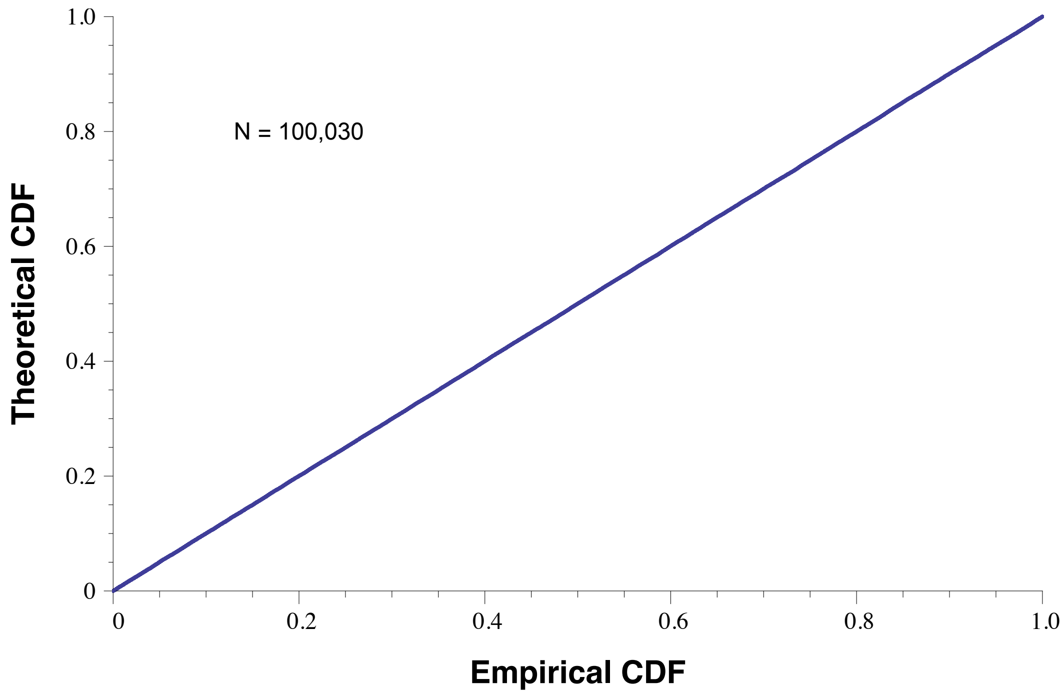

Figure 6: Goodness-of-fit for Model 1

In this Figure, the datapoints are shown as blue dots and the theoretical CDF is shown in gray. The goodness-of-fit seen here is so good that one can hardly detect any difference at all.[10] We can safely conclude that a Gaussian likelihood, common to both models, is unquestionably valid.

5.3 Model Comparison

In this Section, we are looking for a number: the probability that one of our models is better than the other. We hope that this probability is close to one but, whatever it is, it will tell us the relative credibility of the two models and allow us to decide whether one of them is preferable and by how much.

To determine this probability without making any assumptions or approximations, we need to compute the *marginal (global) likelihood* for each model and, from these, the odds favoring one over the other. [9, sect. 3.5][7, sect. 7.4][10, sect. 12.6] The desired probability follows directly from the odds.

[9] termed a probability plot or q-q plot

[10] partly due to the size of the sample

This is a very difficult computation. A marginal likelihood is found by "integrating out" all of the parameters of a model which, in the case of Model 2, means a numerical integration over a space of eleven dimensions, one for each parameter. To make matters worse, numerical integration cannot be done in log space so the software needs to be rather clever. Even with good MCMC software, such functionality is not usually available.

In this study, however, we have values for log(marginal likelihood) so we can carry out this model comparison. The values for Model 1 and Model 2 are -1341.72 and -1317.92, respectively. These values characterize the models as a whole since all of their parameters have now "gone away".

This does not imply that the parameters do not matter; obviously they must. After all, the Weierstrass Theorem guarantees that, with enough parameters, one can always fit the data exactly. However, an increase in the number of parameters[11] will always decrease a marginal likelihood unless this proliferation of parameters actually improves the model. The marginal likelihood is thus a robust measure of model quality.

A larger marginal likelihood means, by definition, a better (more likely correct) model so Model 2 is better than Model 1 but we want this preference to be quantified. Doing the computation in reverse, for clarity, and assuming no prior preference for either model, the odds favoring Model 1 over Model 2, $\mathit{OddsM1}$, is found as follows:

$$\mathit{OddsM1} = \exp(-1341.72 - (-1317.92)) = 4.61 \times 10^{-11} \tag{1}$$

Therefore. the probability that the four instruments are indistinguishable is

$$P(\text{instruments indistinguishable}) = \frac{\mathit{OddsM1}}{1 + \mathit{OddsM1}} = 4.61 \times 10^{-11} \tag{2}$$

which is negligible so we can assert with complete confidence that Model 2 is far better than Model 1.

One-sided results like this are not common and usually occur only if the dataset is very large. Typically, competing marginal likelihoods are closer in value and any conclusion regarding their relative quality is less certain. Here, there is no doubt whatever. These four instruments were *not* equally accurate.

5.4 A Season for Every Time

When you complete an analysis with results that are as convincing as those discussed above, it is easy to feel a great deal of satisfaction. You might even think that the job is done. However, there is always more to be learned. In this example, we have used altitude measurements to assess the errors of Tycho's primary instruments but we have not said anything about *when* those measurements were made. Figure 6 leaves little doubt that the errors were Gaussian but they might still be time-dependent in some fashion. A simple way to check this is with a scatterplot of error versus time.

[11] or the use of a parameter about which the data contain no information

Figure 7 shows scatterplots for each of the instruments. Two of the instruments seem to have little time dependence but errors using the Revolving Quadrant and, especially, the Mural Quadrant are clearly a function of time.

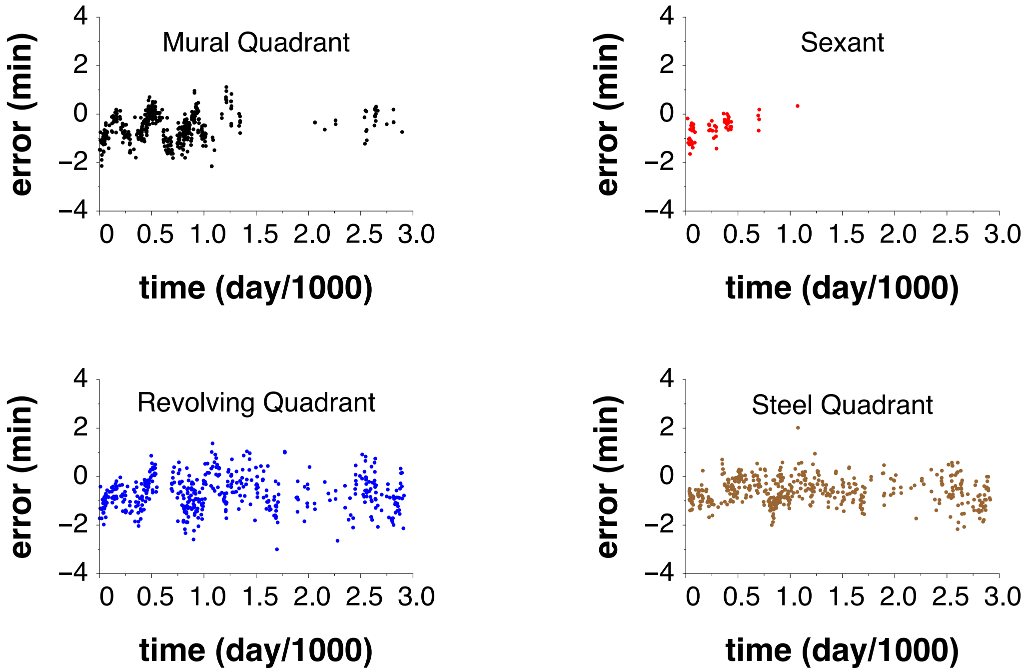

Figure 7: Error Time Sequences

If we examine these plots carefully, it would appear that the errors depend on the season of the year in which the measurements were made. One way to test this hypothesis would be to plot the errors versus something that is definitely seasonal such as the true transit altitude itself. This is done in Figure 8.

In Figure 8, the Sextant and Steel Quadrant errors show little if any altitude dependence while the Revolving Quadrant errors appear to show only a slight dependence, one that is vaguely sinusoidal.

The Mural Quadrant is much more interesting; it shows clearly that its measurement errors got worse whenever the transit altitude was low. One can only speculate as to why this should be so. Does it have to do with the manner in which Tycho measured the altitude of the sun (more easily said than done), an effect due to refraction and/or ambient temperature,[12] some combination of these or, perhaps, something else entirely? It is impossible to say without further information.

Still, it is rather interesting and suggests that there may be even more dependencies waiting to be discovered.

[12]low altitude $\implies$ winter $\implies$ low temperature

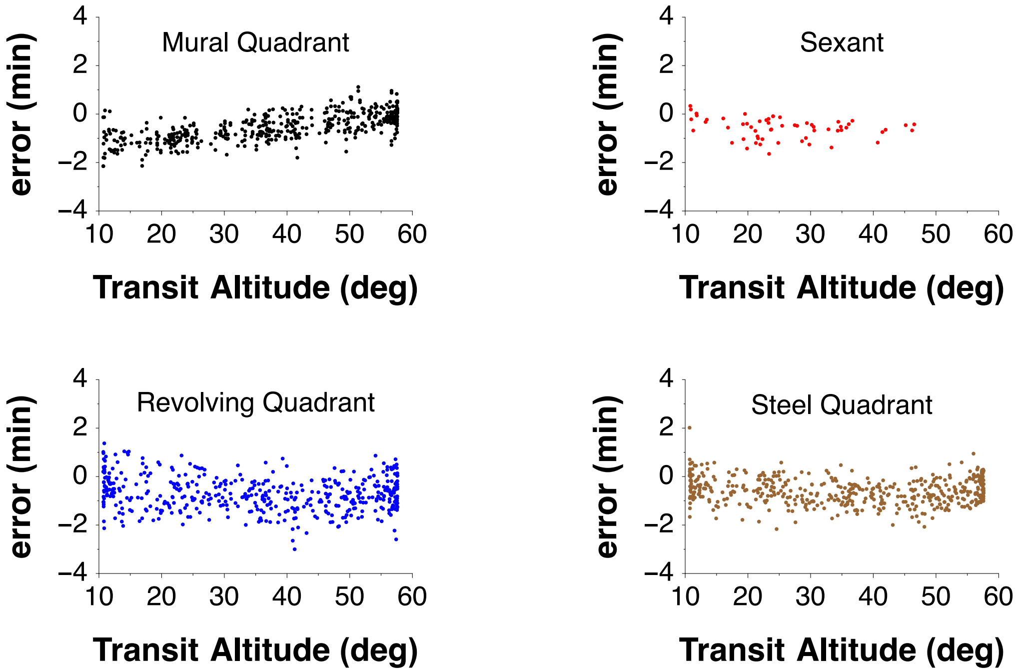

Figure 8: Errors vs. True Transit Altitude

6 Summary

In this study, we have analyzed astronomical measurements made by Tycho Brahe and his assistants to determine how accurate his instruments were and whether they were all equally accurate. We found that all were accurate to better than a minute of arc and that their accuracies were not the same with the Mural Quadrant being the most accurate and the Revolving Quadrant the least accurate.

We did not output just point estimates; we also quantified the uncertainties of these conclusions, finding i) 95-percent credible intervals for the errors of the four instruments and ii) that the odds *against* the proposition that the instruments were equally accurate are more than 20 billion to one.

Finally, we have used this example to illustrate the procedure by which state-of-the-art Bayesian data analysis is done and a few of the reasons why it has largely supplanted the older, frequentist methodology (i.e., statistics). The example discussed was too trivial to exhibit the full power of Bayesian methodology but it does give some hint of the latter.

A English Translation of Figure 2

Observations for the Year 1589

Observations of the Sun

3 January[13]

Meridian altitude of the Sun		
Steel quadrant	12	$42^1/_3$
Revolving quadrant	12	41
Sextant	12	$42^1/_4$ or $^1/_3$
Mural quadrant (both sets of pinnacidia)	12	$41^1/_2$
Declination		
Armillary sphere	21	23 (first)
	21	$23^1/_2$ (alternate pinnacidia)
Replicate declination	21	$23^1/_6$
	21	$23^1/_2$ (alternate)
	21	$23^1/_3$ (middle)

6 January[13]

Hour	$6^1/_2$			clock was verified
Hour	8	6′	40″	upper limb of Sun appears
	8	9	55	middle of sunrise
	8	11	35	complete sunrise

The following afternoon, the clock was running just $8^1/_2$ minutes slow with respect to the $6^1/_2$ A.M. setting.

Meridian altitude of the Sun		
Steel quadrant	13	$14^2/_3$
Revolving quadrant	13	$13^1/_2$
New sextant	13	$14^2/_3$
Declination		
Subterranean armillary sphere	20	51 (first pinnacidia)
	20	$51^1/_4$ (alternate)

[13]Morning observations of solar declinations were determined along with observations of Venus.

7 January[13]

Hour	5			clock was adjusted to Spica in Virgo and the heart of Leo [Regulus] and some weight was added to it
Hour	8	5′	20″	Sun was oriented in the square
	8	10	0	Sun was completely elevated above the horizon

At noon the same day, the clock was running just 5½ minutes slow, an error that crept in since 5 A.M.

Meridian altitude of the Sun			
Revolving quadrant	13	25⅔	
Sextant	13	26½	
Steel quadrant	13	26	25
Mural quadrant (both sets of pinnacidia)	13	25¾	

Declination			
Armillary sphere	20	39½	
	20	39¼	(alternate)

That afternoon, as the Sun was setting in the west, its declination was observed at certain altitudes, as follows, in order to elicit its refraction.

Hour		Solar Altitude		Declination	
2	50′	5	30	20	33½
2	55	5	0	20	32½
3	2	4	16	20	31⅔ (good)

This is sufficiently consistent with the usual observations preceding noon since a daily declination change of three minutes can be realized in six hours.

8 January

Meridian altitude of the Sun		
Steel quadrant	13	38½
Revolving quadrant	13	37
Mural quadrant (both sets of pinnacidia)	13	38 1/12

Declination			
Armillary sphere	20	27½	
	20	28	(alternate)

N.B. Clock is correct at noon.

9 January

Meridian altitude of the Sun			
Steel quadrant	13	50 1/3	
Revolving quadrant	13	49 1/3	
Mural quadrant	13	50 1/3	
Declination			
Armillary sphere	20	15	
	20	15 1/2	(alternate)

There was not enough good weather.

21 January

Meridian altitude of the Sun			
Steel quadrant	16	49	
Sextant	16	48 5/6	
Revolving quadrant	16	48 1/2	
Mural quadrant (both sets of pinnacidia)	16	48 2/3	
Declination			
Armillary sphere	17	16 5/6	
	17	16 3/4	
Alternate observation	16	49	0 (better)
Refraction (subtracted)		6	35
Parallax (added)		2	53
True altitude	16	45	18
Declination	17	19	57
Longitude of the Sun	11	43 1/3	Aquarius
Our ephemerides	11	45	

As luck would have it, there was not enough good weather, or too little or too much subtracted for refraction. The following observation is more certain.

References

[1] BROOKS, S., ET AL., Eds. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC, 2001.

[2] DREYER, I., Ed. *Tychonis Brahe Dani Opera Omnia*, vol. II. Hauniae In Libraria Gyldendaliana, 1915.

[3] DREYER, I., Ed. *Tychonis Brahe Dani Opera Omnia*, vol. V. Hauniae In Libraria Gyldendaliana, 1923.

[4] DREYER, I., Ed. *Tychonis Brahe Dani Opera Omnia*, vol. X–XIII. Hauniae In Libraria Gyldendaliana, 1923–1926.

[5] DREYER, I., Ed. *Tychonis Brahe Dani Opera Omnia*, vol. XI. Hauniae In Libraria Gyldendaliana, 1924.

[6] EFRON, B., AND MORRIS, C. Stein's paradox in statistics. *Scientific American 236*, 5 (1977), 119–127.

[7] GELMAN, A., ET AL. *Bayesian Data Analysis*, 3rd ed. Chapman & Hall/CRC, 2013.

[8] GILKS, W. R., RICHARDSON, S., AND SPIEGELHALTER, D., Eds. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, 1995.

[9] GREGORY, P. *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, 2010.

[10] MCLAUGHLIN, M. P. *Data, Uncertainty and Inference*. causaScientia.org, 2019.

[11] MEEUS, J. *Astronomical Algorithms*, 2nd ed. Willmann-Bell, Inc., 1998.

[12] SIVIA, D. S., AND SKILLING, J. *Data Analysis: A Bayesian Tutorial*, 2nd ed. Oxford University Press, 2006.

[13] TUPMAN, G. L. A comparison of Tycho Brahe's meridian observations of the Sun with Leverrier's solar tables. *The Observatory xxiii* (1900), 132–135, 165–171.

[14] WESLEY, W. G. The accuracy of Tycho Brahe's instruments. *Journal for the History of Astronomy 9* (1978), 42–53.

[15] WESLEY, W. G. Tycho Brahe's solar observations. *Journal for the History of Astronomy 10* (1979), 96–101.

[16] WIKIMEDIA. http://commons.wikimedia.org.