

デジタル惑星科学データの整理と評価 －画像と地震波データの例－

寺 藺 淳 也 (財)日本宇宙フォーラム
齋 藤 潤 西松建設(株)技術研究所

**Indexing and Evaluation of Digital Planetary Data
- Lunar Image and Earthquake data as the example -**

Jun-ya TERAZONO (Japan Space Forum)

Jun SAITO (NISHIMATSU Construction Co., Ltd)

Abstract

近年のデジタル惑星科学データの普及により、研究者がパソコン等を用いて容易に惑星科学的研究ができる環境が整いつつある。また、これらのデータはネットワークを利用すれば常に最新のものを入手できるようになっており、データだけでなく、解析手法やデータについての説明なども、ネットワーク上で全て得ることが可能になりつつある。

しかし、デジタル情報は大量であり、現状で必要な情報を迅速に得ることが困難になるケースが多い。今後、日本においてLUNAR-A、MUSES-C、SELENEといった月・惑星探査計画が始まり、大量の高精度データが得られた際に、それを整理し、研究者が迅速に情報に得られるようなシステムを開発しておくことは、研究の進捗の上でもきわめて重要であると考えられる。そのためには、研究者が利用しやすい、解析に耐え得る良質のデータを抽出・整理・評価する必要がある。

本講演では、月の画像及び地震波データ及びを題材に、抽出・整理・評価するためにどのような作業を行えばよいか検討した結果を報告し、併せて将来的な方向性についても議論を行う。現在、そのための第1段階のアプローチとして、画像についてはクレメンタイン探査機のデータを題材に、画質を表現するパラメータを抽出し、それらを実際に評価する試みを進めている。

また、アポロ計画により得られた月の地震波データは、既に一部をデータベースしている。このデータベースは現在、実際の解析へ応用されているほか、今後はこのフォーマットを基礎として、多種類のデータを1つのファイルに納め、自由に取り出せるフォーマット(やそのためのアクセスライブラリ)の開発、様々なプラットフォームに適応できる環境の開発などを進めている。

こういった試みを組み合わせることにより、月・惑星研究のための「支援システム」を作成できるのではないかと考えられる。

あるべき惑星情報システムの姿

データの概観を把握できると同時に、データの全ての部分にもアクセスできる。

情報へのアクセス方法が統一されていること

データの概観を把握できると同時に、データの全ての部分にもアクセスできる。

より速い解析

概要データと詳細データの両方が用意されていること

研究効率の向上

操作性に優れ、研究者が全てを把握できること

研究者の相互協力の促進

データだけではなく、解析手法やソフトウェアのアップデートなども含む

最新の情報が常に得られること

規模が小さく、モジュール化されていること、また、ドキュメントが整備されていることが必要

...そのためには何が必要か？

いま、我々の研究で進めつつある部分は白字で、これから進めていきたい部分は黄字で表している。

- ・データの内容(画質、データ量など)を表すための指標作り
- ・指標のインデクス化

情報へのアクセス方法が統一されていること

- ・アクセス方法を統一するためのライブラリ・ルーチン
- ・汎用的かつ実用的なデータフォーマット

より速い解析

研究効率の向上

研究者の相互協力の促進

操作性に優れ、研究者が全てを把握できること

概要データと詳細データの両方が用意されていること

最新の情報が常に得られること

- ・GUI化など、扱いやすいシステムの開発
- ・既存の解析アルゴリズムなどの整理

- ・ネットワークに対応したデータベースシステムの開発
- ・最新情報の収集

画質パラメータの抽出

画質評価の作業は、宇宙科学研究所の固体惑星グループが所有するコンピュータ、及びCD-ROMチェンジャを使用して行われた。評価に使用した画像は、Clementineミッションにおいて得られた画像を収めた、Clementine raw image collection 全88枚である。

このCD-ROMにおいては、画像はPDS (Planetary Data System)のversion 3形式で収められている。

この画像ファイルは、先頭に画像情報を表すヘッダ、そのあとに実際の画像データが並ぶ形になっている。

本CD-ROM全てに含まれる画像ヘッダから、画像についての基本的な情報、画質を表すと思われる情報、及びPhotogeologyの観点から重要であると思われる次の情報を抽出した。

画像の中心緯度 (CENTER_LATITUDEヘッダ)、中心経度(CENTER_LONGITUDEヘッダ)

標準偏差(STANDARD_DEVIATIONヘッダ)

圧縮率 (ENCODING_COMPRESSION_RATIOヘッダ)

輝度値の最小値(MINIMUMヘッダ)、最大値(MAXIMUMヘッダ)

入射角(INCIDENCE_ANGLEヘッダ)、位相角(PHASE_ANGLEヘッダ)、射出角(EMISSION_ANGLEヘッダ)

画像ゲイン(GAIN_MODE_IDヘッダ)

なお、ヘッダからの情報抽出を行うプログラムはPerlスクリプトとなっている。

さらに、ファイル及びヘッダの情報を加工し、次のデータも作成した。

ファイルの大きさ (ヘッダ及び実データを含む、ファイル自体のサイズ)

色差 (MAXIMUMヘッダとMINIMUMヘッダの数値の差)

ヘッダ切り出しの作業は、UNIXのシェルスクリプトなどを使用している。また、各CD-ROMを読み出すためのCD-ROMオートチェンジャプログラムは、C言語を使用して作成している。

このように、本作業に使用したシステムは、大量の情報を処理し、かつ今後の変更の際にも柔軟な対応ができる形になっている。

画質の評価基準

今回は、Clementineの画像について、緯度別、ゲイン別での評価を行った。評価の指標には、色差と標準偏差を使用している。ここでは、色差とは、画像内に含まれる輝度値(DN値)のうち最大のもものと最小のものとの差であると定義する。一方、標準偏差は、ある画像に含まれる輝度値の分布の度合を表した値である。

色差が小さく、標準偏差も小さい画像

色の分布は極めて狭い範囲に偏っていると考えられる。従って、画像のダイナミックレンジは乏しく、画質が高いとは考えられない。

色差が大きく、標準偏差は小さい画像

極端に輝度が高いピクセルと輝度が極端に低いピクセルがある。輝度値の広がり小さいため、分布を広げる画像処理(ストレッチング)を行うことにより、画質を高められる可能性がある。

色差が小さく、標準偏差は大きい画像

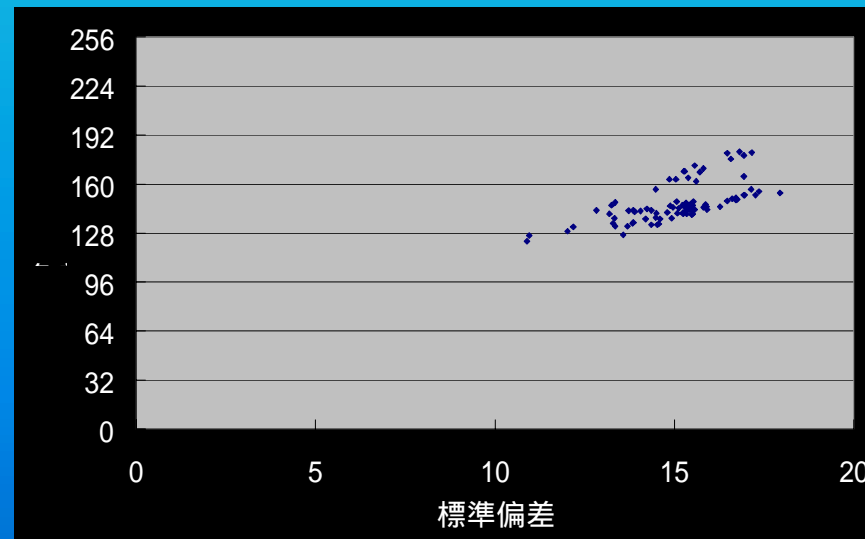
画像としてはいろいろな輝度値を適当な割合で含むと考えられ、画質のよい画像である可能性がある。但しあまりに色差が小さい場合には、ほぼ同じような輝度値となっている恐れもある。

色差も大きく、標準偏差も大きい画像

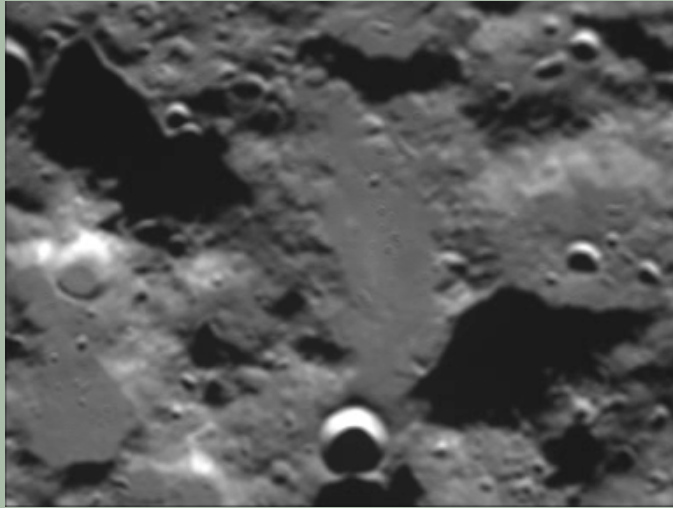
極端な値の輝度値が含まれる可能性もあるが、画質は高い可能性がある。月の場合では、海と高地のような異なる輝度値傾向を持つ領域を撮影したものである可能性がある。

色差と標準偏差の関係

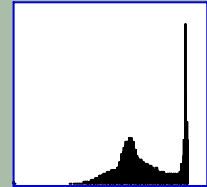
- 全CD-ROMについて色差と標準偏差の平均をCD-ROM単位で算出し、それをプロットしたのが下の図である。Clementineの画像は8ビット(256階調)で表されているから、256以上の色差はあり得ない。
- これを見ると、クレメンタインの画像は2つの系統に分かれることがわかる。すなわち、色差は比較的大きいグループと、色差が小さいグループである。前者はPre-mapping Phaseのように画質の調整が比較的うまくいっていなかったものと考えられる。
- Clementine全体で見ると、色差は128あるが、標準偏差は15程度と、色差に比べるとかなり小さい。このことから、Clementineの画像は全体として、極端な値を含むものの、全体としては階調に乏しいという傾向がわかる。



色差と標準偏差： Example images



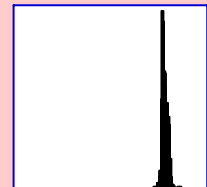
LUA0202B.187
色差： 238
標準偏差： 85.855
GAIN_MODE_ID = "2"



INCIDENCE_ANGLE = 78.71 <deg>
PHASE_ANGLE = 78.65 <deg>
EMISSION_ANGLE = 0.38 <deg>

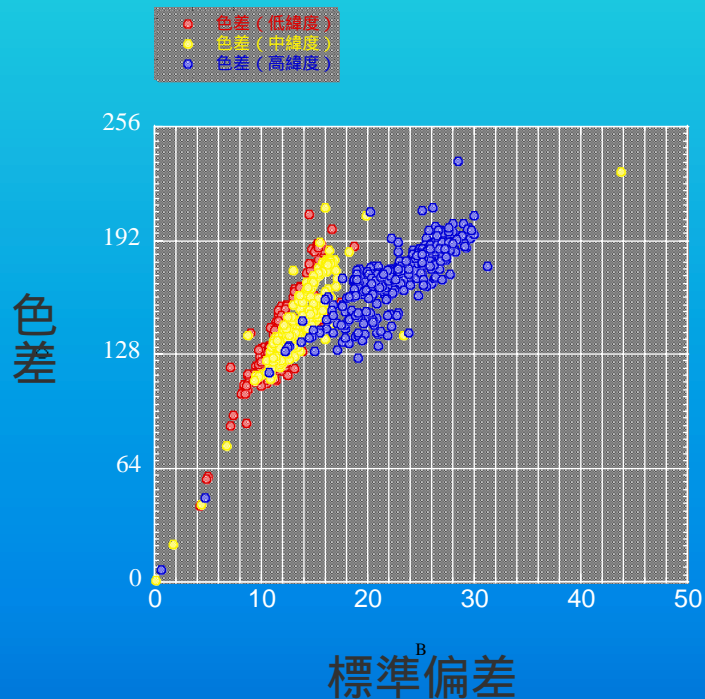


LUA2249I.187
色差： 70
標準偏差： 4.594
GAIN_MODE_ID = "1"



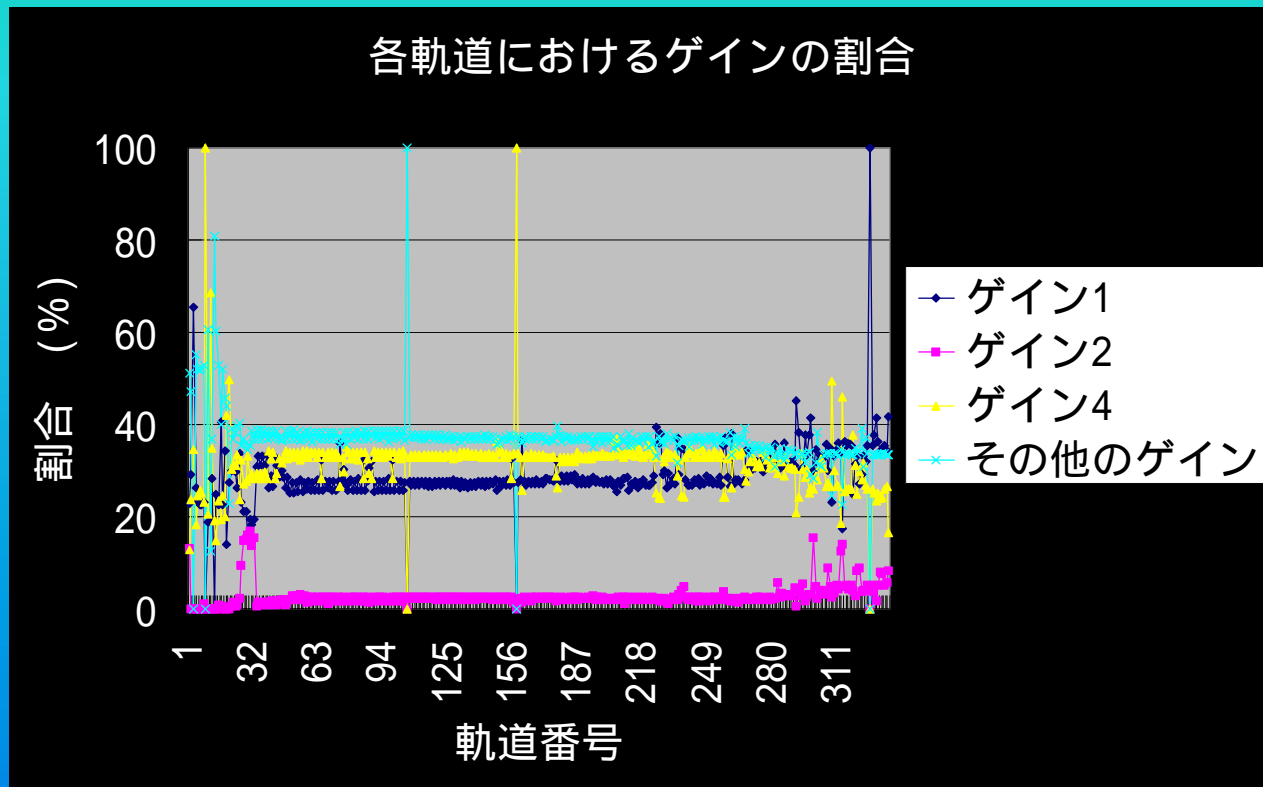
INCIDENCE_ANGLE = 4.26 <deg>
PHASE_ANGLE = 4.51 <deg>
EMISSION_ANGLE = 0.27 <deg>

画質評価ダイアグラム(1)緯度別プロット



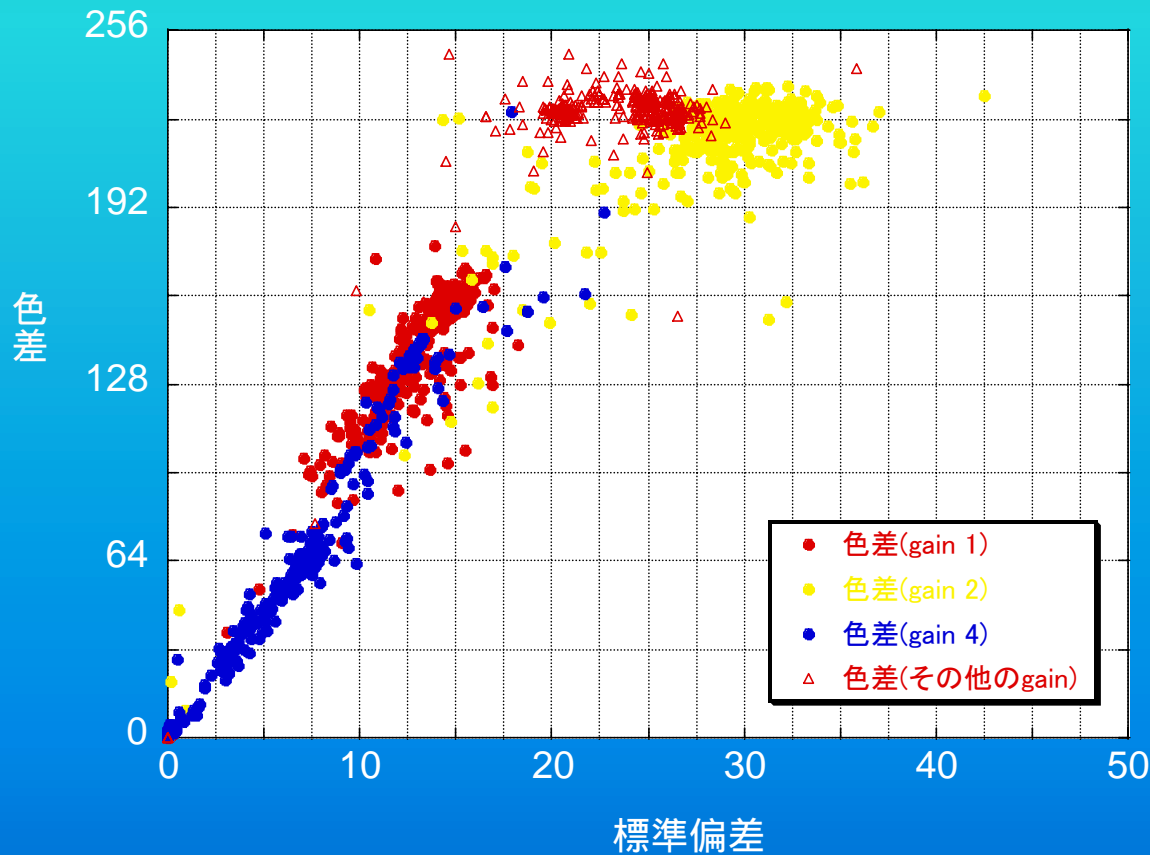
- すべてのClementine画像から、軌道別にデータを取り出し、それを画像の中心の緯度ごとに「低緯度」(0~30度)、「中緯度」(30~60度)、「高緯度」(60~90度)と分類した上で、標準偏差と色差の関係をプロットした図である。
- この図で見ると、低緯度・中緯度地域に比べ、高緯度地域の分布がずれていることがわかる。ずれ方は標準偏差がより大きくなる方向になっており、またばらつきが大きくなる傾向にある。このことは、同じ色差であっても明暗がかなり極端になっている画像が多いことを意味している。
- 低緯度と中緯度の画像を比べてみると、中緯度地域の画像の方が標準偏差がより大きくなる傾向がある。これらの画像についても高緯度地域と同じことがいえるが、中緯度地域の画像は高緯度ほど極端なばらつきを示していないため、比較的バランスが取れたが質のよい画像が多いことを示唆している。
- 低緯度地域の分布は中緯度地域画像の分布とかなり重なっており、似たような経緯を示しているように見える。しかし実際には色差128、標準偏差10付近に固まっている傾向がある。この事から見て、低緯度地域の画像は色調の差が乏しく、従って地質などの解析には必ずしも適さないものが多い。

各軌道におけるゲインの割合



各軌道毎のゲインの割合をプロットした図である。Systematic mappingのフェーズにおいては各ゲインの割合はほぼ一定しているが、Pre-mapping(1~31)及びPost-mapping(295~348)のフェーズでは、ゲインの割合は一定していない。

画質評価ダイアグラム(2) ゲイン別プロット



Clementineでは、カメラのゲインは主に1、2、4という3つのモードを使用していた。特に、解析によく使用されているUVVISでは、ほとんどが1、2及び4のデータである。

左図は、標準偏差と色差のダイアグラムを、ゲインにより分類してプロットしたものである。ゲインは、1、2、4及びその他に分けてある。

1及び4のゲインのデータは比較的標準偏差と色差の分布が広範囲に分布しており、光学的にバラエティに富んだデータが得られることが期待できる。

一方、ゲイン2及び分類外のゲインのデータは、分布が偏っているほか、色差が極端に大きなデータとなっている。これは、輝度値の分布が極端なデータが、これらのゲインの画像に多いことを示唆している。

画像の解析を行う際、このようなゲインのデータからの画質評価の視点も重要である。

illumination geometryのバリエーションのある データセットの重要性

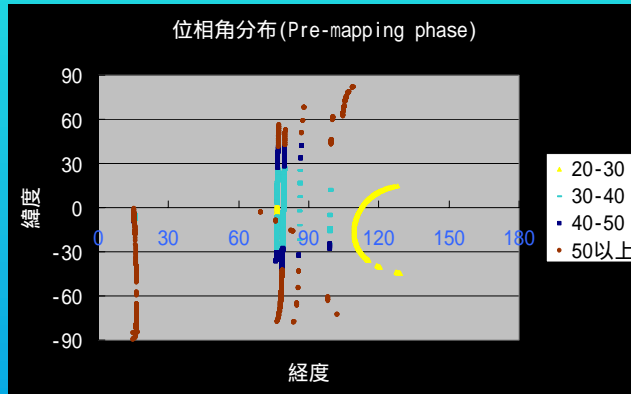
I SISで過去に推奨されていたパラメータの検証から、*photometric function*のパラメータによっては鉄量推定などの情報取得に大きな影響がある事が分かった。パラメータは適時改定されており、常に最新の物を用いて補正式を再設定する事が不可欠である。

ま た、「最新」のパラメータにしてもそれがどこまで適用が可能なかを吟味する必要がある。例えばここでシミュレートした最新の*photometric correction*では*albedo*の違いによる*photometric function*の変化については考慮されていない(McEwen et al., 1998)。

つ まり物質種による*photometric function*への影響(≒*albedo*による影響)まで考慮してより詳細な*correction*を行うためにはuserがそれぞれの物質種の地域について再度決定することが必要となる。

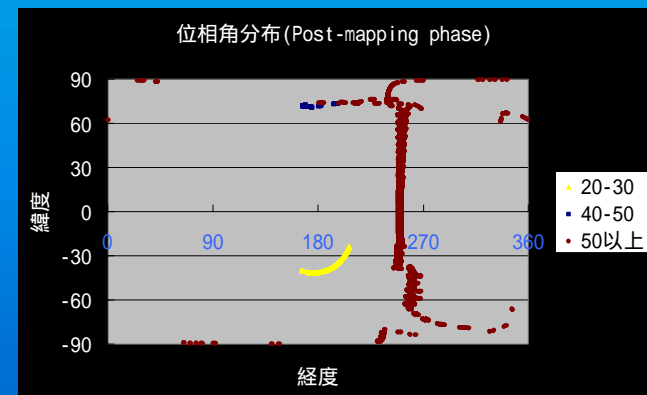
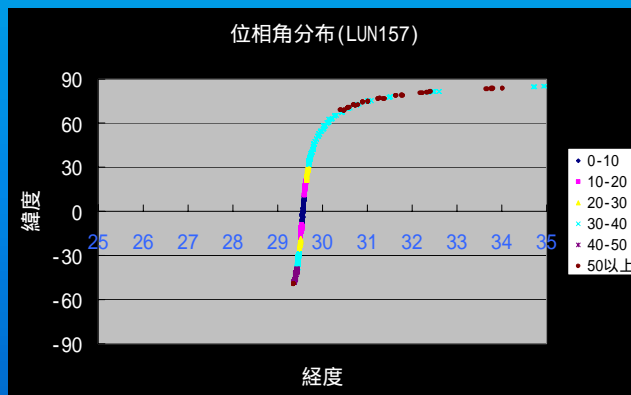
そ のためには異なった*albedo*ユニットについてそれぞれの*photometric function*のパラメータを決定するのに足るだけのバリエーションを持ったデータセットが存在するかどうかを検証する事が必要となる。

Clementine UV/VIS imageでの illumination geometry分布状況



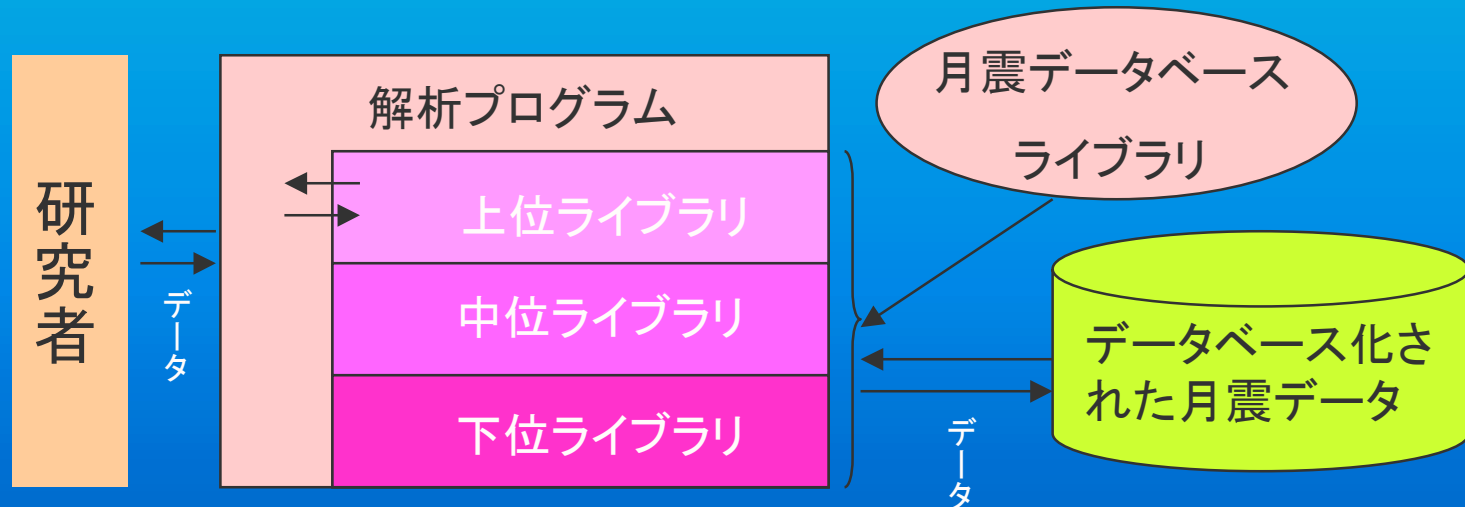
Clementineの軌道のうち、Pre-mapping phase、Post-mapping phase及びSystematic mapping phaseの3つの軌道を選んで、位相角の分布をプロットした。Pre-mappingは001~031(全軌道)、Post-mappingは345~348、Systematic mappingは157を選んだ。

Systematic mappingに比べ、Pre-/Post-mapping phaseでは位相角のバリエーションは小さい。さらに複数の角度での観察を行っている地域が偏在しており、Photometric functionのデータをClementineのデータだけで決定することが難しい。今後、光学的な面から地質解析を行っていく場合には、これらのデータがどのように分布しているかをグローバルにプロットすることが不可欠である。



月震データベースとは？

- ◆ アポロ計画により取得された月震データは、生データにして約80GBにも及び、そのままでは研究者の手に負えない。また、データはそのままでは解析には使いにくい形となっている。そのため、データを解析に適する形に整理し、これらの膨大な情報を効率的にかつ高速に扱えるよう、データ構造を工夫したデータベースを作成した。
- ◆ また、研究者が解析用に作成するプログラムからデータを簡単に、また統一的に扱えるようにするため、専用のアクセスライブラリを用意した。



月震データベースの特徴

◆ 主なデータをヘッダ化

発震時刻や観測地点など、解析に頻繁に利用される情報は全てヘッダとしてある。このため、研究者はヘッダをみることによって必要とされる情報にすぐにアクセスできる。

◆ ネットワーク対応のデータフォーマット

ネットワーク間でデータをやりとりすることを想定し、テキストファイルのフォーマットと、バイナリデータをネットワーク間でやりとりする規格(XDR)を利用したフォーマットの両方をサポートしている。また、フォーマット自体はきわめてシンプルであり、地震波に限らず、画像や他のデータなども、最小限のライブラリの変更でサポートすることができる。

◆ アクセス用ライブラリを用意

研究者がデータ構造の詳細などを気にすることなく、解析プログラムを作成することもできる。ライブラリはC、FORTRANがサポートされ、将来的にはPerl、Javaからも利用可能になる予定である。また、このライブラリを使ったユーティリティ・プログラム(波形表示やプリンタ出力、データ変換など)が用意されている。

月震データベースの現状

◆ LUNAR-Aに向けて、整備活動を再開

一部解析には既に利用されているが、より使いやすく、より広いプラットフォームに対応するための整備作業を再開した。

◆ Event Tape・連続データのデータベース化

これまでデータベース化されていたのは、月震データのほんの一部(Special Event Tape)であった。今後の解析に備え、Event Tape及び連続データについてのデータベース化を検討している。

◆ 一部不足していた情報の整理

データ内に存在するタイムギャップやRise timeなど、まだインデクス化されていないデータについてもインデクスにできるように検討をはじめている。

◆ 月震の統計的分析

月震データベースで整理されている情報を利用して、月震を統計的に解析する試みをはじめている。例えば、発震時刻やそのパターンなどにみられる規則性がどの程度正確なのかを調べることで、月震の発震機構などを推定する手がかりが得られる可能性がある。

2つの「プロジェクト」が目指すもの

◆ データの 量 の問題

将来、惑星探査が一般的となったときに、今あるよりも大量のデータが地上に送られてくる。研究者がその中から必要とするデータを見つけ出し、解析を行う際に、いかにそれを迅速に行うかという問題が今後重要になると思われる。

◆ データの 質 の問題

解析に耐え得るデータを迅速に選び出し、適切なアルゴリズムを適用して解析を行うためには、良質なデータを判定し、それを選び出せるようなアルゴリズムとシステムが必要である。

◆ 研究者が 本当に必要なデータとは？

これまでは研究者が使うデータについて特に注意が払われてきたということはない。しかし、大量のデータが得られる時代を前に、これからは新しいデータ解析ポリシーを確立していく必要がある。

まとめと今後の方針

- ◆ 月の画像と地震波(月震)データを例に、デジタル惑星科学データの整理と評価について述べた。
 - 画像については、色差と標準偏差のプロットによって画質評価を試みた結果、
 - ・ 緯度毎の違いでは、高緯度地域は明暗が極端な画像が多く、また低緯度地域は色調の差に乏しい画像が多いことが確かめられた。
 - ・ ゲイン値による分類を行った結果、色差、標準偏差とゲイン値とはある程度の関係が認められる。ゲイン値も画質評価のパラメータとして利用できる可能性がある。
 - 同一地域を多くの異なったillumination geometryで撮像したペアは少なく、多くの地質ユニット毎にそれぞれのalbedoに応じたphotometric functionのパラメータを設定する事は難しい。
 - 月震についてはデータベースを構築し、その評価をはじめている。
- ◆ 今後の作業方針
 - 今回使用したパラメータ以外にも、他の画質評価の指標についても評価する。例えば、画像のファイルサイズ、輝度値の平均値、圧縮率など。また、画質が良好な領域で、できるだけ多くのgeometryで撮像された領域の共通部分をグローバルにプロットする事で明確にし、これらの地域についてサイエンス and/or 開発利用の見地で探査を設定する事が出来るか否かを調査する。
 - 月震データベースは本格活用を目指し、特に統計的に解析できるところから友好に利用していく。
 - これらのデジタル惑星科学データを有効利用するために、ツール、アルゴリズム、評価基準など総合的な検討をこれからも進めていく。