

紫字は世話人でメモした zoom での質疑を転載したものです。

星の明るさ（等級）較正、統計誤差と系統誤差の現在（2020年版）：鈴木尚孝（Kavli IPMU、東京大学）

野上(京大)大変面白い講演ありがとうございました。黒体放射でぴったり合う天体があるというのが非常に不思議なのですが、どういう天体なのでしょう？

鈴木：野上さま、これは、SDSS百万個の星スペクトルから吸収線がないという条件だけで17個弱見つけたのですが、不思議なことに、温度がみんな一緒、GAIAで距離がわかっているものに関しては、大きさが地球半径くらいで一緒なんです。ヘリウム白色矮星の温度が低いものだと考えられています。水素白色矮星よりは温度が高く水素ラインが消え、ヘリウム白色矮星よりは温度が低くヘリウムラインが見えないという解釈です。理論家の方が温度の低いヘリウム白色矮星という線で、解析論文書いてくれました。Color vs Color diagramのFig 1 & 2に一直線にならんでいます。DC白色矮星という吸収線のない星もあるのですが、DCは一直線にはならないところが違います。

諸隈(東京大)：Hayes 1985の絶対calibrationのVega等級は $\sec(z)=0$ での値になっているのでしょうか？(論文を読めばわかるはずですが...) なっているとしたらどうやって行ったのでしょうか？むかし普通にやっていた「異なる z (elevation)で何度か観測して外挿する」というやり方でしょうか？

鈴木：諸隈さま、現在は、STScI Vega $V=0.03$ になるようになっていきます。実は、STScIのCALSPECのVega Spectrum, STScI Ralph Bohlinさんは2007年くらいまで、 $V=0.00$ でCalibrateしていました。(Susana Deustuaさんと私が指摘するまで)彼は、Vegaが0等級になるようにVegaを定義するものと思っていたのです。が、6個の星の平均なので、Vega $V=0.03$ が定義なのです。現在は0.03が入っています。

等級は、大気外で測定した場合になるように定義されています。理想的には、大気モデルをオゾン、エアゾル、水蒸気、 $\sec(z)$ に分解して、forward modeling で合うようにできるようになればいいのですが。。LSSTは0.9mの較正専用望遠鏡を隣に作るので、これが可能になります。Mauna Keaにも共同で一個較正専用望遠鏡がほしいですね。。。

大塚(京大せいめい望遠鏡)：興味深い公演ありがとうございました。過去の文献を精読することは重要ですね。黒体放射でぴったり合う天体というのは、実際にはさまざまなmetal absorptionがあるが、そのマルチバンドフラックスプロットはBBフィットで近似できるということなのでしょう？これらの星はmetalが多いとか？

鈴木：大塚さま、コメントありがとうございました。実は、Keck/LRISで S/N 100 のスペクトルを撮ったのですが、metal lineは全く見えません。上記野上さんのところで引用した理論論文にあるように、温度の低いDBヘリウム白色矮星のようですが、彼らの論文よく見るとモデルスペクトルよりもBBの方がよく合っています。Magellan、HST、ZTF、DESIでは、

新しい装置がつくと、BB星を観測してくれています。System Throughputが直接測定できるので、便利だと思います。是非せいめい望遠鏡でも観測してみてください。

高見(台湾中央研究院)：大変ためになる講演をどうもありがとうございます。

ところでこちらの分野では、ベガ等級をずっと使い続けています。(50年にもわたる測光結果を、文献をreferしつつ比較しなければならないので、仕方ありません。)

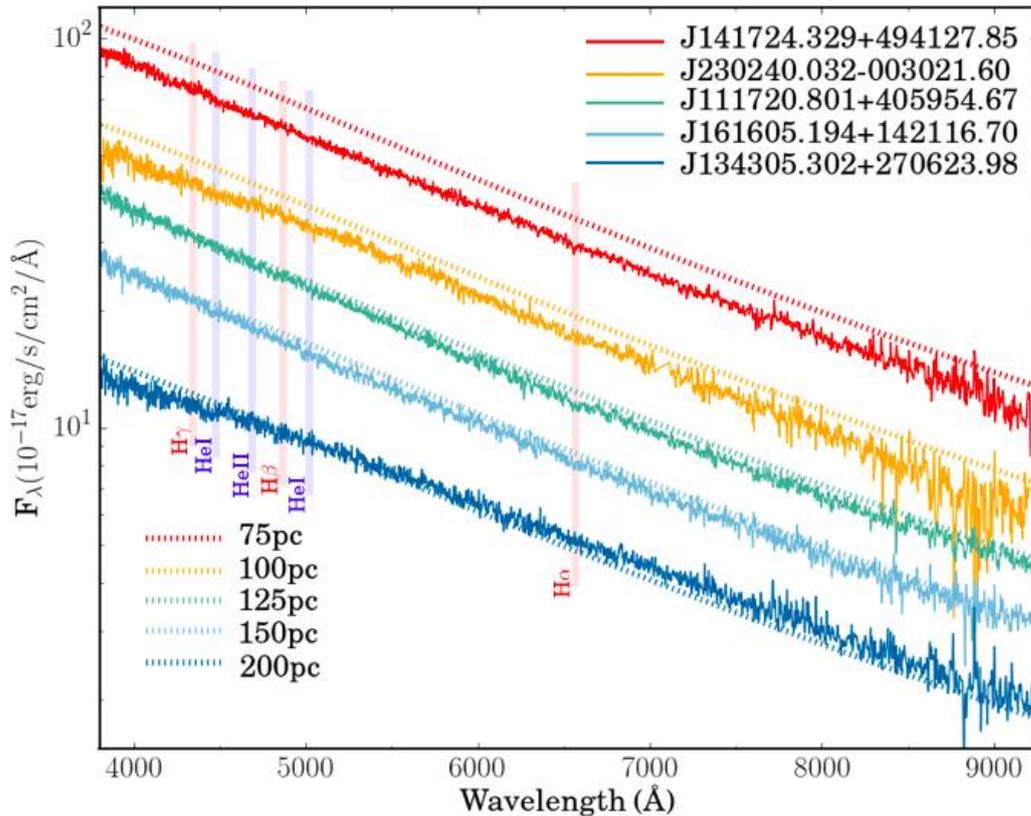
ベガの明るさにも変化があるとのことですが、例えば過去の文献で見つけたターゲット天体の等級と、現在の等級を比較する際に、どの程度の系統誤差がありそうか、もしご存じであれば教えていただければ幸いです。どうぞよろしくお願いします。

鈴木：高見さま、コメントありがとうございます。ベガが変光しているのではなくて、HSTの測定が動いてしまっていると思っています。バックアップスライドにあったのですが、可視光はCCD、赤外はMgCdTeとdeviceが替わってしまっており、Vegaにはパッシェンジャンプがあるので、うまく繋がっていないのが現状だと思います。また、地上では絶対値が測定できないので、Vegaのモデル(実際には正確なものが存在しない)を Hayes 1985の $V=0.03$ でNormalizeして、赤外に延長しているのが現状です。これはどの論文にも出ていないのですが、どのモデルも合わないので、Ralph Bohlinさんは、2つのモデルを手で繋げたいらしいです。実は、Vegaのまわりにはダストのデブリがあって、赤外にExcessがあるのではというのが現在の解釈です。系統誤差は5%あっても不思議ではないです。ちなみに、HST CALSPECは90年代から現在まで比べると、数%動いています。

>どうもありがとうございました。(高見)

家：ヘリウムWDとの解釈とのことですが、4686 HeII も見えないのですね？ 近赤の10830 HeI の観測例はあるでしょうか？ 以下説明ありがとうございます。

鈴木：家さま、Keck/LRIS R=2000, S/N=100 でも見えません。縦軸Logにしてプロットすると以下の図のようになります。水素、ヘリウムの位置を示していますが、見えません。。。これ傾きがみんな同じってことは、温度が全く同じなんです。論文に載ったかっただけですが、福来先生に却下されました。現在紫外領域のスペクトル、赤外のスペクトルをHST WFC3で撮っています。ParallaxもGAIAよりも精度よく測定しているので、半径が0.5%以下の精度で求められる予定です。



古澤(国立天文台): 教育的なトークありがとうございました。20年前の勉強を思い出しました。置いておいて、キャリブレーションの追及という重要なお話だと思いましたが、もう一方、画像・天体測定側で目的のSEDに対して相対測光精度を追い込む手法(フラットフィールドやubercalなど)の今後のトレンド・重要課題についてコメントありますでしょうか。monochromatic flatとか検討は10年前からあると思いますが、今後どうなっていくのでしょうか。

鈴木: 古澤さま、実はHSCについてはまだできることがたくさんあります。DESやPan-StarrsはCALSPECから相対測光をはじめましたが、HSCはPan-Starrsから始めることになっていますよね。是非HSCでもCALSPEC星を観測してみるとよいと思います。現在、External Collaboratorのフランスチームが再解析を進めていますが、non-linearityも改善が期待されます。が、フラットランプの光度の不定性をPhoto Diodeでモニターする必要があります。(このモニター装置、実は買ってあるのですが、使う機会に恵まれていません) もちろんパイプラインチームの較正はとてもよく銀河の研究には何も問題はないと思いますが、超新星の測光には、いくつか試験したいことがあり、フランスチームと進めています。z-band, Y-bandでは、CCDの底で跳ね返った電子がハローのように映る現象が波長に依存していたり、鎌田論文で報告されているCCDのコーティングによるCCD内の Intra Color Sensitivityも較正したいと思っています。

古澤: アイデアありがとうございます。はい、まさにこのようなHWからデータ較正への反映を日本、特に装置チームの周辺にいる我々が主体となってきちんとやっていかなければ

ばと思っています。下とも関連しますが光量のモニタは観測所の装置チームも検討していますね。

鈴木：トレンドとしては、各望遠鏡に専用の較正装置が作られています。Pan-Starrs, DES, LSSTでは、monochromatic flat fieldがあり、レーザーで精確に望遠鏡のthroughputが測定されています。また、フィルターのthroughputも時間変動することが知られており、これらを測定する装置が用意されています。LSSTでは、フィルターのthroughput波長モデルが1 Angstromの精度で測定することが求められています。（これを達成しないと彼らが目指す測光精度がでない）

古澤：HSCでもバンドパスモニタは導入準備はされていて導入されたらぜひ較正に反映できればと思います。専用較正望遠鏡や単波長フラットについてはないものねだりになってしまっていますが、HSCでもぜひ出来るところから一つずつやっていきたいところです。

鈴木：SDSS(2.4m)望遠鏡には、較正用の0.5mの専用望遠鏡がありました。LSSTには、0.9mの較正専用望遠鏡がLSSTのすぐ横に建設されています。スペクトルがわかっている標準星を常にモニターし、オゾン、水蒸気、エアゾルを含む大気モデルを解き、望遠鏡のthroughput、Mirrorの反射率をすべてforward modelingで解く方向です。TMTがMauna Keaに建設することになった暁には、Mauna Keaのすべての望遠鏡チームで協力して、1mクラスの望遠鏡を較正専用望遠鏡として、常時モニターできるようにして、LSST同様、その日、その時の解がforward modelingで解けるような時代がくればいいなと思います。これは、10年先として、とりあえず、Twilight Timeでよいので、お時間のあるときにCALSPEC標準星、Blackbody StarをHSCで撮っていただくと、throughputを知ることができます。欲をいえば、FoVのいろいろな場所、各CCDで撮ると理想的です。

古澤：どうもありがとうございます。BB starsは興味深いのでぜひ検討したいですね。引き続き議論よろしくをお願いします。

森鼻(名古屋大)：10830 HeIが見えていないとのことですが、このような天体の近赤外線2 μ 帯での観測はあつたりしますでしょうか？2.058 μ mのHeIも見えないのかな、ということが気になりまして。もしご存知でしたらお教えいただけましたらと思います。

鈴木：森鼻さま、コメントありがとうございます。現在ハッブル宇宙望遠鏡WFC3のgrismで分光データを取得中です。200nm-1700nmまで撮れますが、2micronまでは撮れません。自分自身は、ご指摘の2.058 μ mは見たことがありませんが、可視のHeI 587.5nmと強度は同様のはずなので、587.5nmで見えていないということは、2micronのラインも見えないと思います。是非JWSTでテストしたいところです。

淡く広がった天体の撮像データ解析について：菊田智史（筑波大）

秋山(東北大)：HSCで検出された広がった成分が見つかった天体についてKeck/WFCIやVLT/MUSEの面分光の結果などと比較されたことはありますか？

菊田：たとえばUmehata+19 <http://www.sciencemag.org/lookup/doi/10.1126/science.aaw5949> Supplementary materialのFig. S2でMUSEでの結果とS-CamのNBの結果を比較しており、

よく一致することが確認されています。我々のデータとKeck/LRISの比較では、LRISの画像のほうで背景のゆらぎが大きく比較が難しかったです。

秋山：ありがとうございます。論文見てみます。

古澤(国立天文台)：トークどうもありがとうございました。スカイメッシュサイズを目標天体によって変えるのは重要なアプローチだと思いました。2パスで解析されているのかなと思ったのですが、1パス目で、メッシュサイズを誤ると本来広がっている天体が点源になってしまったりもするのかなと思ったのですが、その辺りどう調整されているのでしょうか。

菊田：今回は最初のglobal sky subtractionでのグリッドサイズはデフォルト値で決め打ちでした。それより大きなスケールの放射は引かれてしまって検出できないことになりませんが、そこはトレードオフかと思います。探りたいスケールより大きなスカイメッシュのサイズを使うことが重要になります。

古澤：ありがとうございました

山本(東北)：ご講演いただきありがとうございました。Subaru/HSCとKeckの比較がされているスライド6ページ目について質問がございます。Ly α のピークの位置やその周辺の構造が2つ図で異なっているような印象を受けたのですが、これはどうして生じているのでしょうか？説明を聞き逃していたら申し訳ございません。

菊田：KeckのLya画像で見えているいくつかの明るい点は単に星がうまく引けてないものもあります。我々の画像でもカラースケールの範囲的に目立ってはいませんが明るい星は完璧には引けずマイナスの値になっていたりします。その他のdiffuseな部分の違いはHSCがそういう成分をうまくとらえられていないかKeckの画像の一様性が悪いかのどちらかだと思います(が、どちらか決めるのは難しそうです)。

山本：かしこまりました！詳しくご説明くださりありがとうございました。

宮武：銀河団でもスカイ引きの問題がある、スカイ引きのパッケージなどあるか？

菊田：HSCのパイプラインでパラメータ調整して引き算する必要がある。

補償光学観測で得られたPFSの推定の現状と展望：大野良人(国立天文台)

秋山(東北大)：PSF推定へのインプットのリアルタイムデータとして波面センサーのリアルタイムの補正残差データなどは用いられているのでしょうか？そのようなデータを残しておくのは有用でしょうか？

大野(ハワイ)：PSF再構成の場合は、波面センサーデータや可変形鏡のコマンドデータを用いて補償光学の各誤差を決定していますので、できる限りサイエンスデータが取得されている間の波面センサーデータなどは残しておいたほうが良いと思います。例えば、(昔聞いた話ですが)ELTのHAMMONIというAO装置では、今ところすべてのテレメトリーデータを保存することにしたと聞いたことがあります。かなりのデータ量になると思いますが。。

深川(NAOJ)：(1)今の推定方法で一番誤差として効いているのはどのインプット情報なのでしょうか。(2)過去のアーカイブデータを再解析したいと思ったときは、AOガイド星のデータが残っていない限りはPSF再構成は難しいという理解で正しいでしょうか。最後にコメン

トですが、PSFのピーク付近は当然ながら、裾野の部分まで形状が正確になっていると、広がった天体に対しては大変ありがたい話だと思います。

大野(ハワイ)：(1) どれが一番であるという確定的なことは言えないですが、現実の画像だとキャリブレーションエラー（例えば装置起因の収差やNon-common-path-error（波面センサーとサイエンス装置の光路の違いの収差））などの影響が大きい気がするので、それらをしっかりと調べる、キャリブレーションすることがポイントかなとは思いますが。(2) 再構成したい場合はそのときの波面センサーの測定値などが残っていないと厳しいと思います。ある程度パラメータがわかればシミュレーションで再現することもできるかもしれませんが、すべてを再現することはなかなか難しいです。(裾野部分について) 現在の研究はPSFのコア成分で(FWHMやSR)で評価されているので、今後は裾野部分での推定精度の評価も必要ですね。どのくらいの半径まで必要なのか、どのような指標で評価するべきか、要求される精度は？といったところはサイエンス側からfeedbackをいただけるとありがたいです。

高コントラスト撮像におけるAOデータ解析：鶴山太智 (Caltech)

秋山(東北大)：Forward-modeling について、具体的にどのようなことを行うのかを教えてくださいませんか？

鶴山 (Caltech)：惑星のimagingだけであれば、検出された位置にnegative fake sourceを埋め込んで、再度データ解析を行い、最終的にそのシグナルが消える事を確認します(posit ion angleを変えてpositive sourceを埋め込む手法もあります)。こちらだと、単純に主星に対するコントラストからfake sourceの明るさをある程度予測すればできますし、post-processing由来のflux lossが計算できるのでデータ解析のcalibrationにも使えます。ただ最近の観測装置は面分光になっているものがあり、こちらだと主星と惑星のカラーを考慮に入れて、fake sourceの(波長に対する)モデルを一度作った上でforward modelingを行います。

円盤になると、広がった構造になるのでintrinsicな円盤モデル(SED, ALMA観測の情報等から)をまず作成した上で、実際に得られた(unsaturated) PSFでconvolution→その後データ解析を再度行う、という流れになります。ただこちらは非対称構造や微細構造を完全に再現するのは難しいので、ある程度大きな構造のみを再現するというのが現状のやり方になります。

装置それぞれでヘッダー情報やスペックが異なるので、各々でパイプラインを自作しているところが多いのですが、ある程度自分で設定してpost-processing → forward modelingを行うことのできる公開パイプラインもあります(e.g. pyKLIP; Wang, J. et al. 2015 - <https://pyklip.readthedocs.io/en/latest/index.html>)。

秋山：ありがとうございます。モデルする上ではターゲットのPSFの情報に加えて、装置側の性能評価(PSFの視野内依存性や分光特性)も重要になりそうですね。

多波長画像解析におけるSource deblendingについて：馬渡健 (東京大)

秋山(東北大)：例2に挙げられていたベイジアンアプローチというのは左側の方法(TFIT?)とどの様に異なるのでしょうか？フラックスを求めるためにモデルをフィットするという意味では同じことになっているのでしょうか？

あと、例3の超解像の画像を構成するというのは、天体の形状に対して何らかの仮定をいれてデコンボリューションするというのでしょうか？

馬渡 (東京大) : 例2のベイジアンアプローチに関しては、天体位置にprior情報を使ってフラックスを求めるためにスケールされたPSFモデルでフィットする点ではTFITと同じと考えてOKだと思います。一方でフラックス不定性に関して、TFITなど単純なmaximum-likelihood法に対してベイジアン手法の方がメリットがあるそうです。例3の機械学習を使った超解像度画像の構成に関しては、直接的に何か仮定を引いている訳ではありません。ただトレーニングデータセットとして分解能が悪い天体画像と一緒に高分解能画像(正解)も与えてやって、正解を再現できるような生成機(generator)を組んでいます。そのトレーニングデータセットの天体のlight profileが広義には仮定になっているとは言えます。トレーニングデータから大きく外れるような天体データに対しては正しく機能しなさそうに思えます。

秋山 : ありがとうございます、HSTの大量のデータがあるのも効いていそうですね。

古澤(国立天文台) : ピークを引き算していくときにPSFを畳みこむ手法が多いと思いますが、カーネルの追い込みはどのくらい重要なのでしょうか。単純なガウシアンでの代用ではやはり難しいのでしょうか。既存・将来見込みのある手法でPSF・カーネルの標準的な扱い方がありますか？

馬渡(東京大) : どのくらい難しいか一概には言えないと思いますが、自身の経験でSpitzer IRAC画像に対してgaussianカーネルで差引きを行なったときはやはり裾野が全く合わなくて失敗しました。標準的なやり方として、現在SSPでもデータリリースされているような時系列・領域ごとのPSF画像を使用する方法が今後もやられるのかなと思います。実際の観測データから点源を選び出してスタックしてPSFを作る代わりに、実験室での装置のPSF画像を使ってもIRACの場合は結構うまく行きました。

古澤 : どうもありがとうございます。よく分かりました。

竹内(名古屋大) : ご回答ありがとうございます。さらにセッション中の質問の続きです。Confusion問題のもっとも厄介な点は検出限界を下回る多数のソースがある場合のノイズで、これが天体の重心を大幅にずらしてしまうことも多々あります(SCUBA時代の論文にはこれでおかしな結論になったものも多かったですし、遠赤外では現在進行形で難しい問題ですね)。これについては私も昔から研究していますがあまりいい解決法がない状態ですが、今回ご紹介いただいた方法のうまい使い方はないのでしょうか？

馬渡(東京大) : 検出限界以下のresolveされない天体がノイズとなってresolveされてる天体の重心をずらしてしまう問題に対しても、priorを使った天体検出が原理的には有効だと思っています。十分深くまで到達している良いprior天体カタログがある場合は、それで天体重心が決まるため、天体重心がずれる恐れはありません。もちろんよりテクニカルには色々難しい問題があると思いますが、最近のHerschel研究ではprior-based techniqueでconfusion limitより十分深いレベルまで天体検出・測光を達成したとしている成果が出てきています。例えばWang, L et al. 2019, A&A, 624, 98は、MIPS24をpriorにしてHerschel confusion limitより10倍くらいところまで天体カタログを作っており、そこから求めたnumber countやSFRDなど統計量はリーズナブルそうです。

竹内(名古屋大) : なるほど、ご回答ありがとうございます。

竹内 : コンフュージョンが聞く場合にこのような手法は有用なのか？

馬渡：シングルバンドでもイテレーションを繰り返すことで、明るいソースに隠れた天体も見ることがある。

データ科学・機械学習的な手法の応用例：植村誠（広島大）

野上(京大) どういう問題の時にはどういうやり方がよい、というのをざっくりまとめてもらうことってできるでしょうか？

植村(広島)：今日紹介した4つに限ると、まずMCMCはモデルパラメータの確率分布を推定するので、あらゆる場合で使えるはずですが、ただし時間がかかるやり方なので、最適な解一点だけわかれば良い、分布は要らない、という時にはやる必要ありません。ガウス過程は色々応用が効きますが「関数の形がわからない」という時には、それでも内挿して何かしら値を誤差付きで予測しないといけない時に向いていると思います。ただしその予測値や誤差で論文の結論を書くのは危険で、あくまで参考程度と捉えた方が良いでしょう。機械判別はやはり自動で判別したい時です。たくさんある判別手法の中でどれが良いかは難しいです。データが大量（ 10^6 とか）あればニューラルネットも良いかもしれません。よくある規模の問題なら「うまくいけばなんでも良い」という考え方もあると思いますが、今日紹介したSMLRは確率の解釈やパラメータの意味が明確なので気に入っています。

諸隈(東京大)：OU processでのFSRQとBL Lacの区別に興味があるのですが、示された表には polarization degreeのmedianに全てチェックがついていました。polarization degreeのデータがない場合にはどうなるのでしょうか？一番右に示されていたAUC(でしたよね?)はどのくらいの数字になるのでしょうか？

植村(広島)：偏光度を使わなくても他の特徴量をたくさん使えばAUCの値は大きくは下がらないこともあります。確認したら AUC=0.865 がありました。これはすっきりしない結果なんですけど、サンプルが数十しかないのと、そもそもBL Lac とFSRQ がすごくくっきり別れる種族でもないことから、しょうがないのかなあ、と思っています。ちなみに、FSRQに比べてBL Lacはいつも偏光度が高い、という傾向になります。

諸隈(東京大)：ありがとうございます。はい、偏光度の違いが分類にかなり効いて(しまつて)いるのかなと思っての質問でした。偏光情報なしでよいと思えば、例えば、「FSRQかBL Lacか」の分類ではなく「(ラフな)peak frequency」という連続量にして、かつサンプルをFermi blazar+SDSS (or ZTF?)データでやってみる価値はあるのかな、と思いました。

中西(NAOJ)：(本原さんの質問へのフォロー) ASTEは「固い」望遠鏡ですので、主鏡に同架した光学望遠鏡(口径10cmです)で電波望遠鏡としてのポインティングを割とよく再現できます(電波ポインティングには副鏡の寄与があるので、光学望遠鏡で見たポインティングオフセットを完全には再現しませんが)。対して45m鏡は「柔らかい」望遠鏡です(ふにゃふにゃ。わざとそうしてある)ので、光学望遠鏡でポインティングオフセットを測定しても、単純には電波ポインティングのオフセット(の補正)に使うのは難しいでしょう。

植村(広島)：情報ありがとうございます。

本原(NAOJ)：そうなんですよね、でもなんとか光軸を測定しつつ光学望遠鏡の画像とあわせてポインティングができれば観測効率は2倍くらいよくなるんだよねあと思いついていました。

中西 (NAOJ) : 45m鏡の光軸そのものは式があるので計算できるのですが、主鏡に載せた光学望遠鏡がどこを向いているか、の計算が面倒ですね。さらに問題になりそうなのが、光学望遠鏡では測定できない成分が(気象条件に依存しますが)ポインティング誤差の支配的要因である可能性です。すなわち「電波の波面」が揺らぐことは単一鏡のポインティング誤差となって現れますが、これは光学望遠鏡では測定できません(水蒸気ラジオメーターが必要)。

本原 (NAOJ) : ううむ、待機乱流のtip-tiltはそんなに効きますかねえ。可視でも1"程度なので、野辺山のビームサイズだとあまり関係ない気がします。やはり、風で主鏡鏡面が煽られて光軸が動くのが大きいのかなという感触なのですが。

中西 (NAOJ) : 風が強い条件(風速 $>5\text{m/s}$)ではそのとおりですね。一方で、風が弱くてもポインティングが決まらない、というシチュエーションがありまして、それは波面揺らぎの寄与が大きいことが分かっています(以前稼働していた電波シーイングモニターの測定から推定した波面揺らぎから計算したポインティング誤差と実測が感覚的には良く合っていました)。結論としては、ポインティング誤差の要因は環境条件に依存して複数存在する、と言えるかと。(光天連シンポなのに電波望遠鏡のポインティング誤差の話ばかりしてる…)

中西 (NAOJ) :、(河合さんの質問へのフォロー) 風向風速から観測中ある時刻でのポインティング誤差が推定できれば、観測後のポストプロセスで観測データに乗ったポインティング誤差を補正できると期待されます。(浮田さんが意図されているのはおそらくはこれ) これはアレイ型の検出器を使った観測で効果があると期待されます。電波観測データは可視光赤外線のように画像に位置のリファレンスとなるような星や天体が(ほとんど)写り込みません。画像そのものを使って位置ずれの補正をするのが難しいので、今回紹介されたような手法で観測中のポインティングオフセットが推定・補正できればそのご利益は大きいと考えられます。

植村(広大) : ありがとうございます。浮田さんとはこの研究はこれで一区切りでお終い、ということになっているのですが、もし需要があるなら、さらに発展させることは可能ですし、どなたか電波の方が主導してアドバイスすることもできます。

河合 : 望遠鏡の風の影響に対して、アクティブ補正は可能か？

植村 : 今は都度補正で十分に使えるものになっている。今の発表は取り終わった後での指向の補正を想定している。

本原 : 真の指向方向はどのように得ているのか？

植村 : ASTEの中に可視光の望遠鏡が同架されているそうです。

すばるの解析ツールの状況や今後の方針 : Tae-Soo Pyo (国立天文台)

古澤(国立天文台) : 質問というよりコメントなのですが、今後サポートしていくコードの開発をどこでどうやって行い、ライセンスをどうするのか(公開のタイミング・外部からのcontributionの受け方)、といったことも考えに含めて計画すると良いと思います。共同利用サポートしていく時に、ライセンスに気を使うのは案外面倒です。astropyなど大きなコミュニティとどう連携していくのか、使っていくのかなども考慮出来ると良いです。

表(ハワイ観測所) : コメントをありがとうございます。確かにそうですね。後で相談させていただきます。また、観測所中でもソフトウェア中心にデータ解析ツールを纏めるグループが必

妻かと思う所です。一つはエリックさんがGINGA開発でSTScIにもパッケージとして含まれていたりするので、それと関連しても話をすると参考になるかと思う所です。Geminiの方はソフトウェア担当の研究者があったりしてDRAGONプロジェクトを進めたりしますので、観測所の中とユーザコミュニティをどのようにつなげるのかについては工夫が必要ですね。

古澤：ありがとうございます。よろしく申し上げます。

秋山(東北大)：Gemini DRAGON というのは何に基づいたソフトウェアですか？

表(ハワイ観測所)：Anaconda/Pythonベースです。

<https://dragons.readthedocs.io/en/stable/index.html> を参照してください。

高見(台湾中央研究院)：講演おつかれさまです。IRCS (エシエル)やHDSの現在のデータ解析ツールは、一般ユーザーも使えるのでしょうか。

田実(ハワイ観測所)：高見さま、実際にHDSでクイックルックとして使っているIRAFのCLスクリプトはハワイ観測所のページ <https://www.naoj.org/Observing/Instruments/HDS/hdsql.html> からダウンロードして各ユーザーのIRAF環境にインストールして使うことができます。一般的な点光源の分光データでしたらこれでHDS IRAF 解析マニュアルにあるのと遜色ない一次解析結果を得ることができます。実際の観測現場ではGUIベースのログエディタからxgterm上のIRAFにコマンドを送ることでinteractiveかつほぼ全自動でクイックルックを行うようにしています (CALの準備はある程度人的入力が必要としますが天体は全自動)。ただ、やはりこれもIRAFからは脱却したいですね。観測所のSA個人のみでは少々荷が重いので、コミュニティでの協力が得られるととてもよいと思います。

>どうもありがとうございました。(高見)

表(ハワイ観測所)：HDSについては、田実さんからお答えがありますので(田実さん、ありがとうございます。) IRCS Echell1についてはIRAFのタスクの説明で独立したものにはなっていません。

>どうもありがとうございました。(高見)

FOCAS IFUの解析ソフトの紹介と今後の課題：尾崎忍夫 (国立天文台)

秋山(東北大)：IRAF (PyRAF)に依存せざるを得ない(AstroPyでは十分でない)コマンドというのはどのようなコマンドか具体例を教えてください。

尾崎 (NAOJ)：資料にも載せましたが、fitcoords, standardといったたすくはインタラクティブにフィットから除去するデータを選択しやすいので、なかなか手放せません。

秋山：わかります、そうですよね。。。みなさんこれからどうするつもりなんだろう。HSC/PSFのようなサーベイ装置は良いですが、小規模汎用装置は問題ですねえ。

秋田谷(広島大)：上の秋山さんの質問とも重なるのですが、特に分光のpython moduleはまだ乱立状態(かつ未発達)でどれに手を出して良いか迷う状況です。もし検討した際の比較情報、その結果PyRAFに落ち着いた経緯などあれば情報提供していただくとありがたいです。

尾崎 (NAOJ)：色々調べた(つもり)のですが、IRAFのfitcoordsやstandardに変わる機能を提供しているモジュールは見当たりませんでした。このシンポで情報が得られるかとおも

ったのですが、皆さん仕方なく IRAF を使い続けて状況の用ですね。どなたかご存じなら、教えてください。

秋田谷(広島大)：同じ多天体分光器として、昔尾崎さんが手がけていた装置でもある、せいめいの KOOLS-IFU とのソフトウェア開発情報の共有・共通化などがあると効率的だと思ったのですが、状況はいかがでしょうか。

尾崎 (NAOJ)：KOOLS-IFU のソフトは松林さんが開発されていて、話をしようと思いつながら出来ていない状況です。

(zoom で既に出した質問です、どなたかご存知なら教えてください。)

秋山(東北大)：分光データの解析コードで、装置のディストーションに合わせて画像をゆがませる方向ではなく、モデルスペクトルを装置の特性に合わせてゆがませてフィットするような形で解析するコードをご存知なら教えてください。。。

大塚(京大せいめい望遠鏡)：歪み補正+波長こうせいがうまくいったりいかなかったりというのは各チャンネルのアーチャ切り出しと関係しているのでは？ fitcoord の fitting は両端のデータポイントが重要なので。

尾崎 (NAOJ)：言葉で説明するのは難しいですが、reidentify で輝線検出するときに 10 ピクセル積分していて、スペクトルの本来の端から数ピクセル外側でも、積分すると輝線があるように見えてしまって、これらのデータ点は fitcoords のときに手で除去しています。なれば機械的な操作なのですが、初心者には面倒な部分と思います。どのようなデータを除去すべきかは分かっているので、自分でコーディングすればできるのですが、デコミッションが決まっている装置に、多くの労力をかけるのは得策ではないと思って、手を動かしていません。

大金(東北大)：Python ベースの宇宙線除去ツールにはご紹介されていた Astrocrappy の他にも lacosmic というパッケージがあるかと思うのですが、Astrocrappy の方を使われているには何か理由があるのか教えていただけますでしょうか。

尾崎 (NAOJ)：最初は LA Cosmic のホームページで紹介のあったパッケージを使っていましたが、非常に遅いのが難点でした。astrocrappy は非常に高速です。おそらくコアの部分で C などで作っているのではないかと思います。

大金(東北大)：なるほど速度が大きく違うのですね。大変参考になる情報をありがとうございました。

可視光・赤外線偏光観測の原理の確認と課題・将来：秋田谷洋 (広島大)

秋田谷(広島大：講演者)より：実際の偏光観測を直感的に捉える例として、「虹」の偏光動画を以下に共有しました。適宜参照ください。

(1) <https://photos.app.goo.gl/zf1t46KYYXDEsLQH8>

(2) <https://photos.app.goo.gl/gFfNaLfCf8F6gQjG7>

Machine Learning in the Era of Multiwavelength Astronomy：竹内努 (名古屋大)

大澤(東京大)：今回のデータでは銀河多様体が二次元に落とし込めるということでしたが、次元を判定する基準のようなものはあるのでしょうか？また、データに複数の population が含まれている場合にどういった出力になるのか気になりました。例えば2枚の超平面がクロスしていた場合にはひとつの多様体として認識されるのでしょうか。

竹内(名古屋大)：学習で発見される多様体の次元は重要なパラメータの寄与から自動的に決まります。今回は紫外線から近赤外線までのデータで特徴空間を構成しましたので、このデータならば2次元で表現できるということですが、かなり違う物理量をさらに加えていくと、独立なパラメータ数は増えていくと思います。

複数のpopulationがある場合は、分離していれば分離した状態で構成されると予想されますが、クロスがある場合は何もしないと仰るとおり1つの多様体として認識される可能性がありますので、工夫が必要かもしれません。

大澤(東京大)：ありがとうございます！仮に近接している場合に分離できるかどうかもおもしろい問題になりそうですね。

竹内(名古屋大)：非常に重要な点だと思います。ご指摘ありがとうございます！

服部公平(統計数理研究所・国立天文台)：今回の研究の前提は、(名前は忘れましたがロシア人の方の2012年の論文であるように)銀河は何かしらのパラメータに対して連続的に変化するということだと思います。この前提に従えば、多次元空間においてより低次元の多様体を探すためにisomap, UMAPが効果的だということは理解できます。一方で、もしblue sequence, red sequenceの銀河が全く異なる種族であるという仮定に立った場合(多次元空間に2つのblobがあるというような状況)では、これらの手法は使えないという理解は正しいでしょうか。

竹内(名古屋大)：Isomap, UMAPとも大局的なトポロジーは保持しますので、完全にdetachした種族が複数存在する場合は種族の塊(blob)数を保持した分離した多様体が発見されます。今回連続的なものが発見されたということから、blue cloudとred sequenceが連続した種族であることが支持されると結論されるわけです。逆に、もし分離した多様体が発見されたならば連続したパラメータで記述するという仮定が正しくないと結論されます。

分類を目的とした、大局的なトポロジーを保持しない(分類したサブクラスを切り離す)方法もあるので、どういう量を定量化したいかという目的によって、多様体学習のアルゴリズムを選ぶことができます。

服部公平(統計数理研究所・国立天文台)：明快な説明をしていただき、ありがとうございます。種族の個数に関する知見も得られるという点は様々な分野で適用できそうで便利ですね。勉強になりました。

竹内(名古屋大)：こちらこそ重要なお質問ありがとうございました。

小野(東大)：興味深いご講演をありがとうございました。1点質問させて下さい。今回、2次元に次元削減した結果が示されていたと思うのですが、出力次元を上げた結果もあるのでしょうか。

竹内(名古屋大)：はい、2次元が出てきたのは結果で、最悪のケースでは13次元(もとの特徴量空間の次元そのもの)ということもありえました。パラメータの寄与率を評価する方法がいくつかあり、そういう解析によって2次元でほぼ十分という結果が得られたものです。イメージとしては、13次元空間に2次元の曲面が浮いているものが出現しました。

小野(東大)：ご回答をありがとうございます。もしよろしければ、「パラメータの寄与率を評価する方法」(2次元でほぼ十分とされた根拠)についてもう少し詳細を教えてくださいませんか。

竹内(名古屋大)：たとえばですが、決定木(decision tree)による評価をしています。これは次元決定に用いたのではないのですが、コントロールパラメータの数を評価することができ、これが2という数を返してきます。

小野(東大)：ご回答をありがとうございます。コントロールパラメータの数を評価するアルゴリズムがピンときていないのですが、何らかの損失関数を定義するのでしょうか。

竹内(名古屋大)：説明がぼんやりしていてすみません。損失関数型ではなく、線型の主成分解析でいう寄与率に対応する量で、全体の形状を決める重要度の順にパラメータを並べ、いくつ目まで入れるとほぼ多様体の形が説明できるのかを評価します。

小野(東大)：なるほど、ご説明いただきありがとうございます。

谷口大輔(東京大)：一般に機械学習はハイパーパラメーター次第で結果が変わるものだと理解しています。この例えばblob数が1つであるという結論や、データの次元が2次元という結論は、どのくらい(手法やハイパーパラメーターに依存せずに)robustだと言えるものでしょうか？

竹内(名古屋大)：上の大澤さん、服部さんへの回答の補足になります。今回用は銀河進化の定量化という目的のため、大局的トポロジーを保持するアルゴリズムに限定して方法を選びました。このため、理想的な場合には種族数の推定が変動したりすることはありません。Isomapの場合は近傍を探す際の探索範囲がハイパーパラメータになりますが、たとえばこれを極端に小さく取ると銀河の数だけ種族が出現します。推定のよさを測る情報量基準のような統計的指標をいくつか評価して結論を出しています。UMAPはこういった欠点を持たないようにするため、代数幾何的な評価を内蔵していて、ノイズに対してもロバストなアルゴリズムです。大局的トポロジーがIsomapとUMAPで変わらなかったことから、結論はパラメータの取り方によらずロバストであると結論できます。

おそらく、病的な分布を持ってこない限りはかなり安心して使えるという感触を持っています。

谷口大輔(東京大)：なるほど、大局的トポロジーを保持してくれるのは大変嬉しいですね。回答ありがとうございます。

竹内(名古屋大)：こちらこそ、重要なご指摘をありがとうございました。今後の議論に反映いたします。

小野寺：星形成率のシーケンスが見えているように見えているが、exp 関数的進化のパラメータが見えているのか？

竹内：星形成史の違いを調べるのはこれから。

画像高速処理と画像高品質化による銀河形態研究：澁谷隆俊(北見工大)

秋山(東北大)：データアーカイブの上で多数天体の画像へのアクセスは課題だと思います。

澁谷(北見工大)：はい、私も強く思います。機械学習研究では教師データを色々と変えてみたり試行錯誤を繰り返すことになるので、教師データの更新の度に、何日間も画像をダウ

ンロードするのは結構大変です。対策として、1) データの低容量化 (x, y, flux の数値情報のみを抽出できるようにするなど)、2) 機械学習研究もデータの近くで行う (1日目 古澤さん資料)、というような仕組みがあると良いと思いました。

秋山(東北大) : HSCのデータに関しては個別天体の画像切り出しにかかる時間はネットワークの回線よりもデータのステージングの時間が効いているように見えます。天体の数よりもどれだけのフレームにアクセスする必要があるか、フレームへのアクセスが良くまとめられているか、の方がダウンロードする時間を決めていそうです。その意味ではデータの近くで解析を行うにしても、ダウンロードして行うにしても、かかる時間はあまり変わらない可能性もありそうです。hsc_dr に聞いてみた方がいいですね。。。

澁谷 : なるほど。であれば、確かに hsc_dr など内部の人に聞いてみた方が良さそうですね。コメント、ありがとうございました。

井上昭雄(早稲田) : 高解像度化に少し関連して、天体検出やdeblendingも機械でできると思うのですが、そういう試みはどうでしょうか？

澁谷 (北見工大) : はい、天体検出や deblending も機械で出来ると思います。それらも機械でやれば、さらに高速になると思いますが、教師データなどの調整が少し必要だと思いますので、今のところ、高解像度化は講演資料にある高品質化処理で、検出は従来方法で、という方法を取ろうと思っています。機械学習による天体検出や deblending も今後検討してみたいと思います。コメント、ありがとうございました。

小野寺仁人 (国立天文台) : 機械学習でstructural parameterを求めた場合は、誤差の評価はどのようになるのでしょうか？大量のサンプル内での分布を見たいので、個別の天体の測定に付随する誤差はあまり気にしてなかったりしますか？

澁谷 (北見工大) : はい、我々も個々の天体の誤差をどのように付けたらいいか悩んでいます。... 我々の興味はパラメータの分布なので、当面は誤差を付けなくても良いかも知れませんが、最終的には誤差を付けるべきだと思っています。この辺りは、エキスパートからコメントをいただきたいと思っています。

小野寺 : ありがとうございます。モンテカルロで求めることはできるかもしれませんが、試行回数 x 2日 (/コア数) かかっちゃいますよね。

澁谷 : そうですね。モンテカルロで誤差は付けられるかな、と思っています。従来方法の、数年の測定時間と比べると、機械学習の測定時間は十分に短いので、試行回数のある程度減らせば現実的な時間で誤差が付けられると思います。コメント、ありがとうございました。

高見(台湾中央研究院) : 興味ぶかいご講演をどうもありがとうございます。分野外の研究者の、ごく基本的な質問です。

データの質や銀河のサイズによって、どの形態か分類しにくいような銀河はあるのでしょうか。そのような場合、そのあいまいさは、何らかのパラメータとして現れるのでしょうか。どうぞよろしくをお願いします。

澁谷（北見工大）：ありがとうございます。データの解像度や天体の S/N が低くなると、基本的に、どの銀河形態でも分類しにくくなります。また、銀河のサイズが小さくても分類が難しくなります。さらに、遠方宇宙になると宇宙論的な減光が効いてきますので、例えば、渦巻銀河の淡い構造が減光され、渦巻銀河を楕円銀河と間違って分類してしまうこともあります。銀河の合体/非合体でも銀河同士の離角が小さくなると誤分類が増えます。機械学習による分類だと、分類確率が低くなったり、と数値として現れてくると思います。

（高見）ご回答をありがとうございます。(1)「分類確率が低くなったり」というのは、テストデータで判断するのですよね。(2) テストデータでない、分類する1個1個の銀河について、これは**%の確率で形態A、**%の確率で形態Bというような情報は出せるのでしょうか。

澁谷：(1) はい、教師データとテストデータで構築した機械学習モデルで判断します。(2) テストデータではない、1個1個の銀河に対しても確率を出せます。機械学習では形態 A or B の分類をする前に、個々の形態の確率に相当する値を計算しているので、それを参照すれば良いのだと思います。

美濃和（ハワイ観測所）：面白そうな研究ですね。うまく質問できませんでしたが、シミュレーションデータだけと、HSTのデータを加えた場合で、結果がどれくらい向上するかのというのが聞きたい事でした。ついでに、画像高品質化というのは、classical な deconvolution とは違うことをやっているのでしょうか？

澁谷（北見工大）：ありがとうございます。モデル銀河だけだと全然上手く測定できません。例えば、 $r_{e,input} - r_{e,output}$ の図で、モデル銀河だけだと、弱い正の相関は見えてくるのですが、大きな系統誤差 ($>100\%$) が出てきます。Hubble のデータを加えると、講演資料にあるように、大体 1対1 の関係に乗るようになります。実際の銀河には Sersic profile だけでは表せられない、粒々な構造や非対称構造があるためだと思います。[1]

画像高品質化の方は、classical な deconvolution に加えて、「天体のフラックスが空間的に滑らかである」「天体のフラックスが非負である」などの制約条件を入れて、高品質画像を推定しています。これらの制約条件を入れることで、classical な deconvolution だけでは見えない、細かい構造（例えば、銀河同士の離角が小さい銀河合体）が見えるようになります。[2]

美濃和（ハワイ観測所）：どうもありがとうございます。よく分かりました。

美濃和：HSTのデータを教師にする必要があるのはなぜか？モデルによる初期学習だけではなぜだめなのか？

渋谷：HSTのデータを正しい画像として使っている。モデルだけでは含まれない複雑な構造を持っている。

Tomoe Gozenとビッグデータと機械学習：酒向重行（東京大）

鈴木（IPMU）： Transientの機械学習の難しいところはTrainingの為に答え（正例）を知っている必要がありますが、これは人間が実際のデータから目視で作っているのでしょうか？それともシミュレーションで作っているのでしょうか？また、前者なら、どれくらいの頻度で更新しているのでしょうか？膨大なデータですね。。。

酒向（東大）：まず、Tomoeの実際の動画/静止画データにシミュレーション天体（超新星や移動天体）を埋め込みます。これをクラシカルな方法(Sextractorとか)で検出させます。埋め込んだ天体と同じ位置を検出したら正解フラグを。埋め込んだ位置と異なっていたら不正解フラグを付加します。この操作を大量に実施することで、ほぼ同数の正解・不正解の訓練データを獲得できます。いかに実際に近いシミュレーションデータを作れるかが鍵となります。最近、訓練データに（シミュレーションではなく）人間がフラグをつけた実際のデータを付加する試みを進めています。数をかき過ぎないので、どれほど効果があるかはわかりません。まだ、訓練データを試行錯誤している段階なので更新頻度という概念はございません。安定運用できるようになってきたら考えたいです。

ディープラーニングを活用した銀河の形態研究の課題：但木謙一（国立天文台）

秋山(東北大)：ディープラーニングのアルゴリズムそのものはどこまでblackboxで使っても良いのでしょうか？ディープラーニングのパラメータの値が猫・犬分類とは違う物が必要かも知れませんが、それは結果を見ながら設定するべき物でしょうか？

但木：中身を理解しないままディープラーニングの技術を使っても結果は出ると思いますが、他人を納得させることが難しいと思います。ネットワークの構造自体は、犬・猫を分類するものをそのまま銀河の形態分類に使ってもうまくいく気がします（試していませんが）。ネットワークの学習（=パラメータの決定）は、訓練サンプルに基づいているので、犬・猫画像と銀河画像では全く別のものになります。

古澤(国立天文台)：トークありがとうございました。論文化はたいへんという話があり含蓄が深いと思いましたが、具体的に科学論文に昇華させるときに注意すべきことはなんでしょう。機械学習の結果を使う時にこの部分の評価は押さえるべきなど。

但木：短期間で成果を挙げなければいけない任期付研究者という立場上、私は頑張っただけで論文化しました。ただ今は論文化をあまり意識せずに、とにかく色々試して失敗して、コミュニティ全体で経験を蓄積することが重要だと思います。

古澤：ありがとうございます。ぜひノウハウが蓄積されてある種、本質でない部分での人柱だったり手探りの泥沼作業を少しでも避けられる方向になると良いと思います。技術drivenで色々試してみることの重要性はまったくその通りだと思います。

地上中間赤外線データの活用への課題：上塚貴史（東京大）

中島康(国立天文台)：SMOKAでは、すばる望遠鏡でのMIMIZUKUの生データを公開しています。ぜひご活用ください。

上塚（東京大学）：アナウンスありがとうございます。解析ツールがご提供できていない状況で簡単には扱えない状況だと思いますが、今後整備公開していきたいと思っています。

秋山(東北大)：講演とあまり関係ないかも知れないですが、最後にあたりでPWVのモニター値の話に触れられていましたが、外部モニターでのPWVの値とデータに見られるスペクトルの吸収線の深さには良い相関があるのでしょうか？(いや、TMTのサイト評価での議論で外部モニターで測定される PWV 値とデータの質の相関の議論があったので気になっていたの
で。)

上塚(東京大学)：ありがとうございます。すばる望遠鏡での観測データが限られているので分光スペクトルと PWV の関係はきちんとは見られていません。20 ミクロン帯は水蒸気量に強く影響されると思いますが、10 ミクロン帯は大気の窓の端（特に短波長端）を除いて影響は比較的弱いのではないかと思います（よほど天気が悪くなければ）。

秋山：ありがとうございます、3-5umあたりの相関も気になっていました。

上塚：3-5 um も水蒸気が悪さをする波長帯がありますので、そこでは影響はあると思います。IRCSの方がそのあたりは詳しいかもしれませんが…。

高見：これまでのデータを活用するにはどれだけの天体が観測されているのか、クイックブックでわかるようにしてほしい。

稲見：ESO METIS とかでは解析に関してどのように想定されるのか？

上塚：データ解析パイプラインが提供されると想定していた。スペースの IPAC のような組織があると良い。

本原(NAOJ)：本筋とは関係ありませんが、keynote+ zoomの問題ですかねえ

<https://discussions.apple.com/thread/251495100>

HIDES-F自動次元化：原川紘季（Subaru）

秋田谷(広島大)：波長較正済みのスペクトルを得るにはwspectextが必要とのことでしたが、FITS header に波長較正に必要な情報は書き込まれているので、それをastropy.io.fitsで読んで、pythonベースで処理するコーディングは可能ではないでしょうか？(現状は多分wspectextの方が楽だと思うのですが)→直接お答えいただきありがとうございました。可能だが開発労力との兼ね合いでとりあえずwspectextを使っている、とのことですね。分光のどの部分を脱IRAF化できるかは、特に分光のpythonツールがまだ充実していない現状で、分光に携わる関係者には興味大きい問題だと思います。今後も情報共有できれば幸いです。

原川（すばる）：高分散分光のpythonツールについてはIRDでも重要な課題なので、チームと協力して広い用途で使えるようなツールを作成できればと考えています（特にapallあたりの代替が必要）

古澤(国立天文台)：聞き逃したかもしれないのですが、自動処理した結果の基本的な品質情報をデータベースなどに書き出されているのでしょうか。HIDES-Fの生データの公開整備はSMOKAと装置チームの皆さんで進んできていますので、処理済みデータから得られる品質やスペクトル情報もぜひそれらとリンクできるように公開されるとユーザにも有用なのではないかと思います。(昨日ご紹介したアンケートでも、恒星の高分散分光のアーカイブ需要は高いものがあります)自動解析することで、装置の異常検知できるというのは重要な要素だと思います。

原川(すばる)：550nm付近の一次元化後のSNR (photon count)についてはquick lookとして自動観測システム上のwebログ(MySQL)に記録されています。が、fits headerには記録されていません。このあたりは必要な場合には一次元化後にheaderに記録することはできます。アーカイブ及びデータ公開ポリシーについては有志の運用という特殊な状況下ではあるので、今後策定されていくものかと思っています。

古澤：どうもありがとうございます。DBのみでも例えば天文台にダンプリストアで運ぶなどしてSMOKAなどとうまくリンクが取れると有用そうです。検討が進むといいなと思います。

田実(ハワイ観測所)：HDSの場合はシームレスに観測波長が変わると、クラシカル観測での即時データチェックが必要なので少し趣きが変わりますが、似たようなことをしていて非常に興味深く聞かせていただきました。古澤さんもおっしゃってますが、少なくとも一次元化前の解析済みスペクトルについては自動解析でもほぼ一意なものが出てくるので、そうしたものを公開アーカイブできると高分散分光スペクトルはより広くの人に有効利用されるのではないかと思います。もちろんヘッダ等で解析手続きの追跡ができるとういことです。IR AFからの脱却はやはり難しいですかね。特にエシエル分光器はHIDESだろうがHDSだろうが解析手続きは大きくかわりませんので、できれば光データをpythonベースで取り扱うための情報を私も共有して、ツールとしても共通なものを作れば、とは思っています。

原川(すばる)：一次元化前の解析済みデータというのは一次元化直前のデータということでしょうか。

田実(ハワイ観測所)：エシエルのオーダーをくっつけるという意味の完全一次元化、です。ここはcontinuum ベースとするのかフラックス解析するのか、でHDSでは少なくとも変わってきてしまいます(Blazeが安定しないという裏事情があるのですが)。

原川(すばる)：あ、スペクトル一本化のことでしたか。確かに自動解析ではフラックス補正や一本化は少々荷が重いので、そこはユーザーにお願いする必要はありますが、そこまでのデータ処理については共通のものができると思います。

原川(すばる)：みなさん、有意義なコメントどうもありがとうございました。今後も継続的に情報交換などできればと思います。

観測的宇宙論における画像解析の課題：宮武広直(名古屋大)

秋山(東北大)：HSCにおけるPSFの評価のスライドの図の縦軸はどのような数値でしょうか？

宮武(名古屋大)：星のサイズを σ_{star} 、その位置で再構築したPSFのサイズを σ_{PSF} 、とすると、 $(\sigma_{\text{PSF}} - \sigma_{\text{star}}) / \sigma_{\text{star}}$ です。サイズの測定にはHSCのカatalogに入っているSDSS momentsを使っています。見せた図では星はPSFの決定に使われていな

い星を使っています。PSFの決定に使っている星だけでこの図を作るとrequirementを満たすので、PSF interpolationがうまくいっていないのだと思われます。

古澤：Unrecognized blend ほどのように認識されるのか？

宮武：ピーク以外の形状の情報を用いてHSCの画像でblendを認識させている。

視線速度変動による大質量星連星の探査：須田拓馬（東京工科大）

河合（東工大）中国のLAMOSTで同様の探査が行われていると思われませんが、対抗するための観測戦略はありますか？

須田（東京工科大） ご質問ありがとうございます。LB-1の例がありますね。実は昨年中国で今回の話をしたところ、LAMOSTのチームに「我々はブラックホールを発見したよ」と言われてすでに先を越されていたという経緯があります。残念ながら日本の望遠鏡で対抗するのはかなり難しいと考えています。物量作戦では負けるのでアイディアで勝負するしかないというのが現状です。我々の観測計画では短い観測間隔でhigh S/Nのスペクトルを取るの、今のところLAMOSTやGaiaがカバーしていない領域を探査していることになります。

GPU利用画像高速処理一次パイプライン：庭野聖史（東工大）

庭野：公開しているコード(Eclaire)はMITSuME用のパイプラインそのものではなく、それを元に開発した他の望遠鏡でも使えるような汎用ライブラリです。GPUのない環境でもお試しできるように、Google Colaboratory上で動かせるサンプルプログラムも公開しています。
(<https://github.com/MNiwano/Eclaire>)

中島康(国立天文台) alignmentはどのライブラリを使っていますか？

庭野(講演者)：我々の望遠鏡では、一次処理の前に各画像にWCSを貼っていて、合わせる位置はそれをPyWCSで読み込んで求めています。(Astropyにも同様の機能がありますが、PyWCSの方が速かったので) その後の画像シフトはCuPyの行列演算などの関数を組み合わせて自作しました。

中島康(国立天文台) resamplingについてはライブラリを使っていないという理解で良いですか？

庭野(講演者)：それで構いません。自作なので正しい処理が為されているかの検証はしていて、それは論文をご参照下さい。(http://arxiv.org/abs/2008.11486)

中島康(国立天文台) なるほど。resamplingのライブラリを使うと、cudaにうまくのらない(あまり速度アップできない)と想像しますが、いかがでしょうか？

庭野(講演者)：既存のCUDA向けのresamplingのライブラリを使ったほうが良いのではということでしょうか？

中島康(国立天文台)：CUDA向けresamplingのライブラリなんてあるんですか？知りませんでした。

庭野(講演者)：いえ、私も知らないです。如何せん、直接触ってるのはCuPyであって、CUDAそれ自体はCuPyが上手く隠蔽してくれているので……。質問はresamplingのライブラリをCUDAで使うとパフォーマンスがあまり上がらないかを懸念しているということですね。GPUの

アーキテクチャを活用するには上手く並列化されるようなカーネル関数を書かなければいけないので、既存のライブラリをそのままCUDAに食わせても上手くいかないと思います。ただ、画像処理はGPUの本職の一つであって、天文学以外でも需要がありそうなので、CUDA向けのライブラリは探せばある気もします。

中島康(国立天文台)：「resamplingのライブラリを使うと、cudaに...」は「resamplingのライブラリを使うと、CuPyに...」と書くべきでした。まずはCuPyをちょっと勉強してみます。ありがとうございました。

庭野(講演者)：こちらこそ質問ありがとうございました。

古澤(国立天文台)：素人質問だと思いますが、GPUを積んだ計算機を複数台使ったクラスターとしての画像処理といった比較的簡単に使える・または標準的な技術はあるのでしょうか。

庭野(講演者)：一つのマシンに複数のGPUが積まれている場合は、CuPyだけで容易に実装できますが、複数のマシンに跨って処理をさせるとなるとちょっと分かりません。GPUを使用する部分はCuPyで書けるとは思いますが、複数のマシンに処理を分担させるような仕組みは他のライブラリなどが必要になるのではないのでしょうか。

古澤(国立天文台)：ありがとうございます。まずはCPUやバス幅も強めの1台に複数挿しから使い倒してみるといろいろ限界が分かりそうですね。勉強してみます。

秋山(東北大)：multi-core CPU を用いた並列化とGPUの速度は比較したことがありますか？
どんなGPUを使っていますか？

庭野(講演者)：前々からやろうとは思っていますが、まだ取り組んでいません。使っているGPUはTITAN X (Pascal)です。

大澤(東京大)：紹介された図では“Overscan cut out”をCuPy形式にする前にやっていたと思います。GPUにとって苦手な処理が含まれているのでしょうか。こういった処理はCuPyに渡すより先にやったほうが良いという知見があればコメントいただくとありがたいです。

庭野(講演者)：Overscan cut outと言っているのはただの配列のスライス操作で、これはSIMD演算でも何でもないのでGPU上で行う恩恵は特にありません。それから、何枚もの画像を一度に読み込むとVRAMが結構シビアで、VRAMのアロケーションには意外と時間が掛かるのもあって、できるだけVRAMの使用量を減らしたいという動機があります。このため、オーバースキャンはCPU側で取り除いています。

大澤(東京大)：ありがとうございます。特にVRAMがボトルネックになりうるという話は勉強になりました。

[StarSpotMapping with adaptive parallel tempering](#)：幾田佳(京大)

秋山(東北大)：興味だけの質問ですが、、、4-spotの場合の事後分布図はどの様になるのでしょうか？

幾田(京大)：ご質問ありがとうございます。4 spotの場合、各spotのlight curveへの寄与がトレードオフになります。すると、4 spotのうち3つのspotがlight curveを再現して、1つのspot sizeが0に近づくようなサンプリングがされ、各spotのパラメータが行ったり来たり

推移して、等高線に沿いつつパラメータ空間の至るところにmodeが非常に多くあるような事後分布図になります。

秋山(東北大)：ありがとうございます、言われてみるとそうなりそうですね。

すばる望遠鏡・赤外線視線速度測定装置IRDのデータ解析：観測直後の簡易解析から本格的な視線速度導出までのプロセス：葛原昌幸（アストロバイオロジーセンター）

秋山(東北大)：(すみません、他の講演の内容とごっちゃになっているかもしれませんが、、、) IRDの解析の中でIRAF/PyRAFから切り離せない部分は具体的にどの様なコマンドでしょうか？現状、エシェル分光や多天体分光のデータ解析でインタラクティブな部分がIRAFから切り離せなくて困っているように思います。一方で SDSS / PSF などの解析パイプラインで使えるところもあるかもしれません

＞ 返答（葛原より）： PyRAF もしくはIRAFの機能として利用しているのは、`apall` に関連するものです。残りは、apallを部分的/発展的にに利用している、散乱光の除去や、Flat のnormalizeに関連するものです。この残りの部分は、apallと同様のものがpythonで作成できれば、組み込むのは比較的容易なのではないかと推測しています。apallを利用している理由は主に二つあって、一つは歴史的な理由です。HDSや岡山のHIDESなど、高分散分光のスペクトル抽出の機能は主にirafをベースにしてきています。そのため、それを踏襲するのは、多くの人にとって受け入れやすいと思いました。もう一つの理由は、この部分は、他の部分に新しく作成するのが少し時間がかかりそうだと判断したためです。それでいたら、既に存在している安定な機能を利用するのが、まずは無難だと判断しました。

SDSSのパイプライン、PFSのパイプラインの両方ともよく知らないのですが、module として利用可能でしょうか？ そうでしたら、部分的に組み込ませていただくというのは、良いideaかもしれません。一方、非常に長い時間がかかるわけでもなさそうなので、時間を見つけて自分で開発に取り組むのは、学習の観点や、何か新しい問題点を見つけるためには意義があるかなと思っています。

研究会でも出てきた、pyraf/irafはサポートが終了しているという問題、ブラックボックス化している部分の他にも、(これも出てきたかもしれませんが)メモリー内での処理をすることができないので、わざわざいちいち様々な中間ファイルを書き出さないといけないというのも、現代の環境にはそぐわず、高速化やpythonとのデータの受け渡し、HDの整理や省エネという観点で、不便なことが多いと思っています。

ちなみに、高分散分光のpython-base の解析ソフトとしては以下のものがあるようです。まだ試したことはないです。

<https://arxiv.org/pdf/1609.02279.pdf>

秋山：回答ありがとうございます。SDSS / PFS のパイプラインの apall 的な部分のみをモジュールとして使えるかどうかですが、私もトライしたことは無いのでわかりません。

メモリー内での処理の話は逆の立場もあるかもしれません。中間ファイルが残されていると、途中から解析作業をやり直す事が容易というメリットはあるように思います。HDへのアクセスで解析の時間が決まってくるとすると注意が必要ですが。

＞ 葛原からの返答2回目： 少し言葉足らずでしたがメモリーに関して、途中からやり直すために必要なものや、再現性の検証に必要な中間生成ファイルはHDに残した方が良いと思います。また、中間ファイルを残せば、何が起きていたかも確認しやすくなると思います。もう一つ、パイプラインによる一般的なデータ解析から離れた個性のある解析をする場合は、中間生成ファイルを残すべきだと思います。一方、本当に一時的にしか生成しないデータもあるので、それらを全て残すのは、結構大変だと思いますし、あまり意味もない気がします。ディレクトリーの中があまり意味のないデータでいっぱいになってると、振りかえるのも大変になる気がします。それらの点からもやはり、メモリー内で処理できることが前提で、それを踏まえて必要な中間生成ファイルを残すというのがbetterだと思っています。また、やはりスピードが必要になると、メモリー内で処理できないと結構しんどいと思います。

野津：スペクトル抽出をiraf から抜けるのに難しいのはどのあたりか？

葛原：インタラクティブに処理するところを実装するのが手間がかかっている。

+ irafみたいにoptionが多いものを準備するのは大変そう。

議論：議題のある方は以下にご自由にご記入ください

- 日本としての今後のデータアーカイブやデータ解析ソフト開発の体制
 - データアーカイブとその維持体制
 - ソフトウェア開発の予算規模は適切か？
 - ソフトウェアを作った研究者(特にポスドク、学生)は正しく評価されているのか？(e.g., 待遇, 進路)
 - 天文学者(サイエンティスト)とソフトウェア開発者(エンジニア)の分業？
 - IRAFからどのように足を洗うか
- 各個別プロジェクトごとの解析の課題
 - 多波長データの統合をするには？
 - 例えば PyeIt (半自動データ解析Pythonパッケージ)等、世界のトレンドへの日本からの参加は？
- 数理統計や機械学習からの観点
 - 研究会
 - 例：統数研・池田さん方主催の研究会「天文学におけるデータ科学的方法」
 - 議論中心、ハンズオン系ワークショップも必要？
 - 大学院カリキュラムに必要？

以下は議論のメモおよび後日出したコメント・質問：

葛原：機械学習などを活用してデータハンドリングの面倒な部分を処理できると良い。一方でそのような課題は予算が取りにくい。

古澤：データの品質保証でも同様に活用できると良い。

小野寺：IRAFからPythonに移行するにあたっては長期的なサポート (Python2:3問題など) に不安がある。IDLは昔のバージョンもサポートされている。

大澤：Python 2/3 の切り替えは問題としてある。Python の言語自体がマルチスレッドで動かないのはネックと考えている。5-10年のタイムスケールでPython が主流のままかどうかは不安がある。

古澤：ADASSではあまり次世代の解析環境についての話は出ていなかった。(→ 正確には、出ていないというのは脱IRAFという主旨の話です。Jupyter/K8S/dockerやコードリポジトリの話、FITSの代替としてのHDF5などの話はあります。古澤)

上塚：面分光の研究会でも脱 IRAF が話題になってAFRACとかの議論があったが、その後はどのように進展しているのか？

(関係者不在)

植村：(9/16 10:15) aflak の関係者です。補足します。aflak は computer visualizaion の専門家である藤代先生（慶應大学）との共同研究で、面分光データのより良いハンドリングを目指しています。実際の開発はその研究室の院生さん1,2名が担当していて、githubで公開しています。<https://github.com/aflak-vis/aflak> これまでに空間方向の任意の領域の選択や、波長方向の簡単な演算、輝線の等価幅マップの作成、などが可能になっていて、最近は速度場マップに取り組んでいるところです。このプロジェクトに今、最も必要なのは、ユーザーと、ユーザーからのフィードバック、そして天文学の論文に繋がるような具体的な課題と機能の提案です。可視化の研究者にとって、応用先の研究に実際に役立っているということがとても重要です。天文学者が全く想像してなかったようなツールですので、最初使う時にどうしてもハードルが高いと思いますが、それ故にポテンシャルも大きいと思いますので、興味ある方は関係者にコンタクトをとって頂けると大変嬉しいです。天文側では私や松林さん(京大)が窓口ですが、我々を通す必要はありません。

深川：天文台の中でも解析ソフトウェア・アーカイブの体制の改革について議論がある。今はプロジェクト毎に境界があり、ソフト開発もプロジェクトの期間で切れるのは問題である。ALMAの中ではソフトウェア開発の側面でキャリアアップできる道筋はある。

上塚：大学の装置開発プロジェクトとATCではつながりがある、ソフトウェアでも大学との連携した開発があると良い。

深川：ソフトウェア開発に関わると天文台へのキャリアパスになる点についても議論がある。

上塚：天文台の中にソフトウェア開発のプロがいると助言がもらえる。

中島康(国立天文台)：Pythonでマルチスレッド、マルチコアできますよ。

(秋山：上の大澤さんのコメントの記述は私の聞き取りミスかもしれませんが、確認お願いします>大澤様)

中島康(国立天文台) : Python2->3問題について。将来にまた同様のことが起こるかという
と、Pythonは今後2->3のような非互換なバージョン更新はしないだろうというのが core py
thon-dev ML のメンバーの認識のようです。3.9の次は3.10, 3.11 ... と続くという見方が
有力です。

田中(法政) :

IRAFとは関係なくて、どのツールを使うのが良いのか?という問題でふと思ったこと
が、機械学習やMCMCなど、Python や R で色んなツールがありますが、みなさんがどれを使
っているのか気になります。どれを使うのが良いのか、おすすめがあったら教えて欲しいで
す。

私は、MCMC に関しては、

https://www.amazon.co.jp/dp/4320112423/ref=cm_sw_r_tw_dp_x_i5CyFb36G79PC

を読んで分かりやすかったので、PyStan を、

最近始めたディープラーニングは、とりあえずこの入門書

https://www.amazon.co.jp/dp/4873117585/ref=cm_sw_r_tw_dp_x_R3CyFbPW6Q00R

で使われていたものを使っています。とりあえず実装出来て満足してる程度で、どのツール
が良いのかは全く検討しておりません。

あと解析環境も気になります。私はディープラーニングを勉強するために、無料の Googl
e Colaboratory を使っていますが、みなさんはどのような環境でしょうか?