**PACE SAT Uncertainties Working Group initial questionnaire responses**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Surname** | **Forename** | **Do/will your proposed algorithms propagate uncertainties for each retrieval, and/or do/will you rely on sensitivity analyses or post-launch validation to provide overall estimates of uncertainty?​** | **What do you think are the major gap(s) in your uncertainty budget, or aspects that are hard to quantify​ (i.e. measurement uncertainty, forward/inverse model uncertainty, spatial, temporal, or environmental conditions)?** | **How do/will you communicate uncertainties and general range of applicability to data users, e.g. via “error bars” and/or discrete quality flags? ​And is there a plan to validate these uncertainty estimates, if a validation dataset is available?** | **What are the main statistical metrics you plan to use to evaluate your algorithms’ performance?** |
| Barnes | Brian | Yes, I plan to propagate uncertainties, estimated using sensitivity analyses. The general approach is to forward model Rrs spectra derived for typical water conditions (adding both/either of spectrally correlated and spectrally uncorrelated noise), then feed it back into the inversion and assess optimization results against the known inputs. | Forward/inverse model uncertainties are substantial and difficult to quantify, especially lacking a rigorous validation dataset. I also expect uncertainties to be highly dependent on the level of benthic contribution to the Rrs, (% optically shallow). | My plan is to provide(1) a flags field which contains bitwise descriptions of the inversion results/process – e.g., which, if any, parameters settled on thresholds, whether the solution error is less than a pre-defined threshold, etc.(2) pixel-wise uncertainty data products (“error bars”) for each retrieved parameter. Unfortunately, there is not a good validation dataset for these determinations, so the estimates will be assessed through simulations (see #1) | Median symmetric accuracy, symmetric signed percentage bias. Again, partitioned according to %OS. |
| Frouin | Robert | Uncertainties (bias and standard deviation) will be associated to each retrieval. This will require modeling the measurement, identifying all possible error sources (e.g., noise in the input variables, imperfect or incomplete mathematical model), and determining the combined uncertainty. Uncertainty will be expressed as a function of parameters readily available from applying the retrieval algorithm, i.e., clear sky value and cloud factor (effect of clouds on surface flux) by simulating for a wide range of situations the OCI measurements and the corresponding surface fluxes. Simulations will be performed using the SMART-G Monte Carlo code using multi-year MERRA-2 hourly data as input. The large number of data points will allow adequate sampling of atmospheric variability and in particular many variations of daytime nebulosity, for all latitudes. Noise will be introduced in the simulations of the OCI signal. The noise covariance matrix will be obtained by examining actual OCI imagery with structure functions. Additive varying coefficient models will then be developed to relate uncertainties to clear sky value and cloud factor, where the coefficients of the linear relationship are functions of latitude, longitude, day of the year, and eventually other auxiliary variables. The multivariate functions will be expressed as the sum of univariate functions (additive assumption, otherwise fitting would be difficult and interpretability diminished) and determined using penalized smoothing splines. | No major gap, but simulations will require a lot of computer time. Also, noise covariance matrix will not be known before PACE operational phase, and only algorithm uncertainties, i.e., those due to model approximations and parameter errors assuming that the top-of-atmosphere OCI reflectance is known perfectly, will be available at launch. | Uncertainties will be communicated in the form of maps of bias and standard deviation associated with the radiation product. Evaluation will be accomplished by comparing the uncertainty obtained from simulations (as described above) with the one estimated using match-ups between satellite-derived quantities and measurements at various sites. | R square, bias, mean absolute error, and root mean square error. |
| Hasekamp | Otto | We will provide uncertainties for each retrieval. Using Optimal Estimation. | The most important aspect to quantify is the regularization error, because it depends on the prior covariance matrix which is hard to get a reliable estimate for. Our approach is to use a large ensemble based on aerosol model simulations and perform synthetic retrievals. We then adjust S\_a such (within reasonable limits) that the retrieval error agrees with the difference between retrieved and true properties in a statistical sense. See an example at the end of this document.Further, the approach only accounts for errors that are (pseudo-) random and not for overall biases. | We will provide a quality flag based on filter criteria. Most important will be the chi2 of the retrieval.In addition, we provide an error bar which represents 1-sigma uncertainties. The error bar will only be applicable to valid retrievals. | We plan to evaluate the estimated uncertainty in the way as indicated in the figure at the end of this document. the distrubution of (x\_ret -x\_true)/sigma, where sigma is the standard deviation that follows for the state vector error covariance matrix. This should be a Gaussian with a mean value of 0 and a Full Width at Half Maximum (FWHM) of 2 if the error covariance matrix is computed correctly.For the validation of aerosol properties, we will use Mean Absolute Error (MAE), Root Mean Square error (RMSE) and bias. |
| Hu | Chuanmin | Both. For the former, we rely on 1) algorithm uncertainties from in situ measurements and error propagation theory to propagate uncertainties in reflectance to uncertainties in data products; 2) post-launch data product evaluation using high-resolution measurements (to evaluate lower-resolution data) and field observations. | There are not enough field measurements to understand algorithm uncertainties. It is also inherently difficult to validate post-launch data products because floating macrolagae can be very patching within a pixel and outside a pixel. Using drones to map floating macro algae for at least several pixels of PACE may be the way to go, but it will be very difficult given the limited coverage of drones. | Currently uncertainties are NOT provided at pixel level, but at spatial/temporal bin products after considering all uncertainty budgets. | Comparison between low-resolution and higher-resolution data products.Metrics are R, RMSD, Mean difference (to serve as a bias term), range. |
| van Diedenhoven | Bastiaan | Possibly. If we can calculate it for our LUT-based retrievals. However, measurement noise is generally not the main contribution to uncertainty in cloud retrievals. Observations are generally more ‘noisy’ than expected from measurement noise alone and this ‘noise’ is mainly created by inhomogeneity. Observations in broken cloud observations where different angles may seem different parts of clouds, creating ‘noisy’ data. However, this effect may be limited at the HARP-2/SPEXone pixel size.Accurate uncertainties per pixel for the cloud reff+veff retrievals and ice shapes may be difficult, but uncertainty estimates in a region may be more feasible. The statistics of RMS of the fits in a region may be used to determine a ‘noise’ level of the data in that region, which may be used to estimate uncertainties. I’ve been looking at that with RSP data. -Our project includes quantifying uncertainties for various cloud fields using LES and 1-D and 3-D radiative transfer for most of the products.- We proposed to perform cloud optical depth retrievals from all polarimeter angles separately and explore how this can be used as a metric of inhomogeneity and perhaps uncertainty. | -Since uncertainty is mainly caused by inhomogeneity, determining or quantifying the inhomogeneity of a cloud field is important but difficult.-Aggregation of multi-angular data to the surface instead of cloud top is a source of uncertainty we proposed to quantify using simulations. However, since this depends on the cloud field may be challenging to estimate the uncertainty on a pixel level. | - at least, RMS and/or other metric of model fit to data- Where possible this RMS and/or other metrics could be ‘translated’ to uncertainties of the retrieved quantities. We need to see how robust this is using the simulations and if this can be done in a systematic way accounting for different cloud fields.- Validation with in situ measurements is very difficult for clouds… - I also expect a quality flag or explicit data filtering to be needed. | The cloud retrieval algorithms are simple by design and no optimal estimation or something equivalent is used. A simple semi-empirical ‘translation’ of metrics defining the fit (RMS, correlation, etc.) to variable uncertainties is probably what can be used. |
| Lyapustin | Alexei | SEE SLIDES FROM SUJUNG GO<Products> Atmosphere - CM; WV; AOD; spectral imaginary refractive index (k, b) for smoke and dust;Land surface - Kiso, Kvol, Kgeo – coefficients of RTLS BRDF model (0.47, 0.55, 0.66, 0.87, 1.24, 2.13µm); Hyperspectral SR (BRF); ~80-100 values per pixels for 350-890nm range with 5nm step in atmospheric windowsAOD Uncertainty (over Land) : Initially calculate theoretical AOD uncertainty in response to surface reflectance at blue band for each retrieval at pixel level, to assess surface brightness and guiding the aerosol retrieval algorithm. Surface reflectance is the leading contribution to the total AOD error budget. The AOD uncertainty is calculated using Jacobian (dTOA/dAOD) in the blue band and TOA reflectance uncertainty perturbed with BRDF at given geometry. It will be also reported in the atmospheric properties file (as shown in question 3). BRF Uncertainty (over Land) : We provide the BRF uncertainty (Sigma\_BRFn)in Red and NIR bands deﬁned as a standard deviation of the nadir-normalized surface reflectance (BRFn) over the accumulation period of the queue (4–16 days) under the assumption that the surface is stable or changes linearly in time. This uncertainty may include contribution from gridding, undetected clouds, errors of atmospheric correction including those from the aerosol retrieval, and of surface change when reﬂectance change is nonlinear over the length of the queue (number of collected measurement days). We have plan to do post-launch validation using reference dataset listed below (Table 1). Detailed performance metrics are described in the answer of question 4. Cloud mask (CM) product will be provided with discrete quality flag (see question 3). | SEE SLIDES FROM SUJUNG GOThe Table 1 listed products may include the following errors hard to quantify : Gridding uncertainty including instrument spatial response function – will affect all products except CWV;Model parameter error – joint (AOD, refIM, AAE) retrieval from OCI will be affected by assumption of size distribution (SD) and aerosol vertical profile (H). The error will be quantified but it may be considerable when true SD and H are significantly different from assumed climatology. Multi-angle measurements will provide a good constraint.Forward model error (will be small) including assumptions in RTM and unknown physics truth.Sensor noise (SNR) and systematic sensor degradationProduct errors due to clouds (especially cloud adjacency) are hard to quantify.The theoretical AOD uncertainty neglects other contributions, e.g., from variation in the aerosol model, measurement uncertainty, however, works well except for bright surfaces where AOD uncertainty is estimated to be high (~ 12%). - The uncertainty of UV spectral is harder to quantifying as AERONET currently does not provide UV-based inversions. Diagnostic error of surface reflectance will be additionally quantified around AERONET site locations where AERONET aerosol and CWV will be used for AC of OCI to generate “ground truth”. Will require OCI subsetter for AERONET sites.Full error analysis will include uncertainty and bias (=systematic error). | SEE SLIDES FROM SUJUNG GOMAIAC products will have discrete quality assurance flags (QA) at pixel-level, as well as AOD and BRF uncertainty (also at pixel-level). The following are an example of quality flags of MODIS products. This discrete QA flags are assigned based on the algorithm results. For most applications, we recommend to only use data with QA.AdjacencyMask=Clear (000). To select the best-quality AOD, one should use QA.QA\_AOD=Best\_Quality which combines the best values of cloud and adjacency masks: QA.CloudMask=Clear and QA.AdjacencyMask=Clear.For the best-quality BRF, one should apply the following QA ﬁlter: QA.AODLevel=low (0), QA.AdjacencyMask=Clear, and QA.Algorithm Initialize Status=initialized (0). The QA structure may be changed in the future.In addition, we may output rmse from the joint AOD-refIM-AAE retrieval characterizing joint uncertainty of aerosol retrieval and knowledge of surface. This will help guide user’s decision.For the products listed in Table 1, we can apply expected discrepancy (ED) and calibration skill score (Equation (4) and (5) in Sayer et al. (2020), respectively) to validate the product uncertainty (TBD). | SEE SLIDES FROM SUJUNG GOThe basic metrics of following will be used in validation.Simple diagnostic error equation for products listed in Table 1 can be calculated (TBD).Bias, slope, R, RMSE, mean absolute error, mean error |
| Odermatt | Daniel | This project aims to improve freshwater remote sensing methods by taking into account the vertical variation of water quality parameters. Several datasets including (i) automated high-resolution AOP, IOP and CTD profiles obtained by Thetis profiler from a moored research platform (LeXplore), (ii) AVIRIS-NG (later OCI, if available), (iii) in situ supporting IOP, AOP and CTD data (mainly for calibration, ground-truthing and covering spatial variation), and (iv) operational hydrodynamic simulations and forecasting. See more details [here](https://pace.oceansciences.org/proposals_more.cgi?id=41). We implement sophisticated strategies to quantify the full uncertainty budget of derived products (e.g., vertical non-uniformities of CDOM, CHL, and TSM in two layers and the layer boundary depth). The developed models and algorithms, and consequently the uncertainty estimates, are not finalized yet. However, post-launch validation using in situ measurements will be the primary approach to estimate the uncertainty of the retrieval algorithm. Depending on the form of the final model, we will possibly support the uncertainty estimation by performing a sensitivity analysis with respect to the input parameters, e.g., physical stratification. The “law of propagation of uncertainties” will be employed for this step. But, we must emphasize that, due to the inherent complexity of the involved phenomena, we most likely implement stochastic models, e.g., a machine learning approach, which will make sensitivity analyses a bit difficult.We follow standard approaches for an accurate assessment of individual components of the uncertainty chain. After the definition of a traceability chain, equations and the sources of uncertainty, we use L1 radiance uncertainty estimates of involved sensors (AVIRIS-NG, later OCI) if available. We consider uncertainties introduced by data pre-processing including geometric correction (e.g., radiometric changes due to resampling), atmospheric compensation (e.g., radiometric changes caused by applied and likely uncertain atmospheric transfer functions) estimated via vicarious calibration based on in situ radiometric measurements from the LéXPLORE platform and in situ measurements, i.e., (i) and (iii) in the project summary above. Further, we evaluate uncertainties of information retrieval schemes using benchmarking approaches complemented with in situ observations. | a) Several uncertainties must be considered to facilitate an uncertainty budget, including those related to instrumental effects and calibration, as well as those associated with involved processing steps (i.e., geometric and atmospheric correction, information retrieval). b) Measurement uncertainty, in particular of Thetis profiler, is a source of error. Systematic in situ measurements in the LeXplore region are being performed. They include CHL, TSM, and CDOM water samples at different depths, as well as IOP and AOP measurements using other instruments. Such datasets are used to regularly calibrate the Thetis profiler, and therefore reduce the uncertainty of the final products.c) Uncertainty associated with the spectral response function is also often difficult to be quantified. It depends on spectral, instrumental and environmental noise, which its quantification requires detailed knowledge of the instruments and sophisticated laboratory experiments.d) Although the main source of in situ data in our project is a stationary monitoring site (i.e., LeXplore), spatial variation should also be investigated for the uncertainty estimation of the final products. We plan to conduct a systematic field campaign to tackle this part of the project. AOP, IOP, and CTD profiles are measured. However, it must be emphasized that the vast majority of our data will origin from a littoral site (< 1 km from the shore), which most likely is not applicable for OCI. | The validation datasets include both Thetis data and in situ measurements. In a simple manner, using different instruments/methods, and to a lesser extent performing the field measurements by different people will help to validate the uncertainty evaluation. Furthermore, an optimal processing based on models will be defined and subsequent systematic assessments considering compromised data quality, data resolution, processing assumptions, etc. will enable identifying sources of uncertainties and their impact at the product level. Calculated global and (depending on the input data) pixel-wise uncertainty estimates can be added to the algorithm outputs as a data quality flag. | Different metrics will be used for different deliverable models:a) A model that estimates optical stratification from the physical stratification: Mean Absolute Error (MAE) to check the model accuracy. We will also calculate the coefficient of determination (r2) to measure the goodness of fit of the model. Following the more recent studies, we will also implement some new recommended criteria, such as the slope of the regression lines and performance indexes, to check for the external validation of the model on the testing datasets. The model is considered valid if it satisfies the required conditions. b) A model that relates vertically non-uniform IOP to Rw: We calculate the RMSE, the Percent Difference (PD) and the spectral angle between observed and modeled results to assess the model performance. It will be followed by an in-depth assessment using other error and decision metrics, e.g., intra-pixel Coefficient of Variation (CV). |
| Ottaviani | Matteo | The retrieval scheme is based on inverse methods (specifically, a Levenberg-Marquardt type of inversion). Once measurement uncertainties are provided at input, he method automatically provides at output the uncertainties associated to the retrieval parameters (ocean surface refractive index, AOT at 870, Cox-Munk windspeed). | The forward model is rather accurate for the job. The major sources of uncertainties will derive from the actual sensors’ polarimetric uncertainty, and will grow proportionally to the aerosol load (whose polarization signatures interfere with those of a “clear” sunglint pixel. One caveat deals with the presence of foam that will likely affect our retrieval uncertainty and will be difficult to validate (foam albedo is negligible at the RSP wavelength of 2.2 micrometer used in the concept studies, but not at the HARP2 870 nm wavelength). | The overall algorithm will classify sunglint-contaminated pixels via a flag. The uncertainty on the ocean surface refractive index will be derived from the covariance matrix and given as error bars derived from the covariance matrix. | Direct comparison with the results from other inversion codes, and available products (e.g., aerosol optical depth or sunglint mask).No specified metrics. |
| Pahlevan | Nima | Our goal is to produce aph, bbph, Chla, and SPM products.Our proposed algorithm is a class of NN that learns a mixture of Gaussians. The network directly parameterizes the uncertainty within the learned covariance matrices of the mixture components. The model is trained with in situ data only. Our goal is to be able to produce per-pixel uncertainties using prescribed/modeled uncertainties associated with input features (Rrs). Post-launch validation data will further help with validating our uncertainty products. We should be able to produce preliminary per-pixel uncertainties for our Chla, SPM, and aph by the end of this calendar year. | Measurement uncertainties in the training data (Rrs and IOPs) and how to quantify them are the main concern. | The goal is to produce error bars. QA flags may come handy when AC fails. Yes, we would like to validate our validation procedure.  | Here are three metrics we’d like to use (Morley et al. 2018) – We just renamed MSA and SSPB to Error and Bias to keep it simple. C:\Users\asayer\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\25193A56.tmpC:\Users\asayer\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\B87D0D4.tmp where C:\Users\asayer\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\262E7202.tmp;(Type II) linear regression slope These are technically variations of Seegers et. al. 2018 metrics that are more interpretable. The median can be replaced by mean (if the test data are of high quality) to get a more complete picture of the performance. |
| Stamnes | Snorre | Yes to the first part.No to the second part. Although sensitivity analyses and post-launch validation are both important, we do not rely on them to provide estimates of uncertainty. Sensitivity analyses will be conducted to test the performance of the algorithm for synthetic data, to generate an idea of overall performance of aerosol/ocean properties for the PACE observing system.Also, retrievals using PACE-analog airborne data will be performed, and to validate the results comparisons will be made to HSRL in-water and aerosol measurements and in-situ ocean measurements, as well as AERONET overpasses, see response to Q4. | Requires SPEXone, HARP2 and OCI error models for L1C data, containing systematic and random uncertainties.Ideally requires measurement error covariance matrices including off-diagonal elements for SPEXone, HARP2 and OCI.See attached Knobelspiesse et al., 2012, specifically pp. 21469-21470 (pages 13-14 in the PDF).These matrices may need to be measured pre-flight in the lab and/or updated based on solar/lunar/appropriate vicarious calibration methods.Otherwise we will do our best with what we have to try to construct these matrices properly. I am not sure if these measurement error covariance matrices have ever been measured or provided by any satellite instrument team. We are also open to suggestions here, but the approach outlined in Knobelspiesse et al., 2012 was what we proposed to do in order to obtain the best estimates of the posterior error covariance matrix, in turn leading to the best estimates of the uncertainties.There may be some refinements to the posterior error covariance matrix to remove a priori biases, etc. But those refinements should be straightforward to implement, and would still benefit from a full measurement error covariance matrix. | Posterior error covariance matrix will be provided for corresponding to all retrieval parameters, which directly provide “error bars” for each parameter, the accuracy of which will depend on the measurement error covariance matrix including off-diagonal elements, see above. | We would compute RMSE, R and the least squares bisector for the PACE-MAPP retrieval parameters for the following cases: 1. PACE-MAPP retrievals performed using PACE synthetic data, which is an insufficient but necessary step.
2. PACE-MAPP retrievals performed on PACE-analog airborne data, with comparisons made to HSRL (aerosol+ocean parameters), AERONET (aerosol parameters) and in-situ (ocean) measurements.
 |
| Twardowski | Mike | Uncertainties cannot be directly propagated forward or in the inversion for ZTT. There are also inherent assumptions in any algorithm that will be specific to water type and will be the primary sources of bias error, so propagation of error through the algebraic manipulations of an algorithm may not have much meaning.We are assessing MAPE and bias uncertainties by:1) Bootstrapping uncertainties in input parameters2) Applying to synthetic data sets with presumably negligible inherent error. However, this includes assumptions in construction of synthetic data set that may be inaccurate. The synthetic data set may also include unrealistic cases. IMO, synthetic data sets can be used to provide general guidance in algorithm development but not rigorous quantification of algorithm uncertainties.3) Applying to high quality field data sets where a separate rigorous closure analysis has been done to quantify the inherent errors in the data. IMO, this is ultimately the only way for rigorous error assessments. For in-water inversion algorithms and BRDF effects, etc., performance assessments can/should be done with data now, i.e., we should not wait for post-launch validation. | The largest source of error in algorithm performance assessments is ac9/s absorption measurements. The correction for scattering in the tube is a function of water type, so cannot be reliably quantified. Technology for measuring absorption accurately in-water is a big gap.Inherent errors in data sets, even our highest quality data sets, are larger than the errors in the fundamental radiative transfer approximations used in SAAs. This is not only true for ZTT but also Gordon and Morel relationships. Empirical algorithms like QAA have an advantage in a way in that common bias errors in measurements are included in the empirical data sets as well as the validation/assessment data sets made with the same sensors. Availability of very high quality data sets overall is a gap.We see the largest source of error in GIOP-type approaches (including ZTT inversion nested in GIOP) is the limited number of shape vectors for IOP subcomponents (specific absorption spectra), primarily for phytoplankton absorption. We need a large, representative, very high quality hyperspectral aph data set to develop optimal shape vectors. This is a critical gap to improving performance of GIOP-type algorithms.Multi-angle radiance is very rarely measured in any field work or validation program, which means BRDF effects are very rarely assessed. BRDF corrections with typical nadir viewing radiometers are also never carried out to account for changing relative reflectance as a function of solar zenith. Multi-view radiance measurements should be standard for any validation work. | MAPE or RMSE in reconstructed vs measured Rrs can accompany every inversion result.We are not sufficiently along in our ZTT error assessment analyses and implementation to prescribe quality flags. At some point we expect this will be possible.Validation, i.e., performance assessment for the algorithm, is being done now for several different data sets. Future field validation efforts must be improved in quality and scope to assess errors and how errors get dispersed through algorithms. Validating a SAA for just a single product provides weak insight on algorithm performance and the potential for refinement; data sets allowing for full forward/inverse closure and bootstrapping are the most insightful. | MAPE and bias error.RMSE is also being assessed, but errors are not normally distributed, so this metric is more challenging to interpret. |
| Zhai | Pengwang | MAPOL is a research algorithm based on least squares fitting of the polarimeter data. There are two ways to estimate uncertainties for this type of algorithms. One is based on the Jacobian matrix at the converged state vector, which can be calculated based on radiative transfer model. If efficiently implemented, this can be provided for each retrieval. the other is based on the post-launch validation as you mentioned in the question. This requires co-located independent data source, which is impossible to guarantee uncertainty estimation for each retrieval.For our proposal, we will be more focused on the second approach, i.e., provide uncertainty estimate based on co-located independent data source, for instance, AERONET, HSRL lidar, etc. The reason is that there is doubt in the community on the reliability of the Jacobian approach, which was manifest when our paper went through the peer-review process.Meng Gao: Uncertainties from error propagation using Jacobians are in development in MAPOL (most of the coding part is done, but need extra efforts for evaluation and testing). | The major gaps will be the availability of in-situ measurements so that we cannot do a global uncertainty estimate. One approach could be the comparison of uncertainty estimation from the first and second approach and gauge the accuracy of the uncertainty from the Jacobian approach, which could be used to fill the geographical and temporal gap limited by in-situ measurements.Meng Gao: if the in-situ data is not available for all the retrieval quantities, simulated/synthetic data are also useful to evaluate uncertainties. | I would prefer an error bars type of uncertainties. As I briefed above, we would use co-located in-situ measurements to evaluate the uncertainty of the MAPOL algorithm. | We would use the mean bias and variance of the error distribution. Again, we prefer to provide uncertainties based on independent data source, so it would be less applicable to each individual retrieval. A more meaningful method would be the overall uncertainty of a set of retrievals from a particulate geographical area and temporal period. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |