Afghanistan opium survey report 2024 - Methodology

This document covers various methodological aspects regarding survey design and estimation procedure.

1. Estimation of area under opium poppy cultivation

Remote sensing methodologies have been used by UNODC since 2002 to monitor the extent of opium poppy cultivation in Afghanistan. Changes in the location of opium poppy cultivation and the increased security difficulties involved in accessing the area of interest require continuous improvements of the survey designs.

A sampling approach is used to cover those provinces where most of the poppy is found, whereas a targeted approach is used in provinces with a low level of opium poppy cultivation. “Targeted approach” means that a certain area of a province is fully covered by satellite imagery.

From 2015, new and better satellite technology allowed for a major change in the study design: the size of the grid cells used for acquiring satellite imagery has been reduced from 10 x 10 km images to 5 x 5 km images. This change affected only provinces where a sampling approach was used; all other provinces were not affected by this change.

In 2024, out of 34 provinces in Afghanistan, 17 were sampled, 9 were targeted, and one both, sampled and targeted. The remaining 6 provinces were considered to be poppy-free based on information from the field. The main change has been that Takhar has been moved from Targeted to Sampled, as there had been indications that the amount of poppy cultivation was increasing there. Another important change has been that the sampling of Badakhshan has been increased by 66% to cover all areas in the province given its growing importance in opium cultivation. Also in Badakhshan, images have been acquired twice with a month of difference in order to ensure that poppy sowed in a staggered manner was identified. Smaller changes have been the addition of an additional target area in Nuristan and the subtraction of eight contiguous sample areas in Nimroz.

Other than these changes, the same sampling locations have been used since 2019, which ensures high levels of comparability of the annual estimates.

Table 1 Area estimation method, by province, 2024

Region	Targeted approach	Sampling approach
Central	Kabul, Parwan, Logar, Day-Kundi	Day-Kundi
Eastern	Kapisa, Nuristan	Kunar, Nangarhar, Laghman
Northern	Samangan	Faryab, Jawzjan, Balkh, Sari-Pul,
North-eastern	Baghlan, Kunduz	Badakhshan, Takhar
Southern	Ghazni	
Western	Hirat	Badghis, Farah, Ghor
South-western		Helmand, Kandahar, Uruzgan, Zabul, Nimroz

2. Study design

The study design applied in the Afghanistan opium survey was developed in 2015 and since then adapted for selected provinces when appropriate.

2.1 Sampling frame

The sampling frame was established by extracting the area of land potentially available for opium poppy cultivation in 18 provinces. This area was divided into regular 5 km by 5 km grids, which constituted the sampling frame. The final sampling frame, from which the satellite images were randomly selected, consisted of 8,208 cells. In the case of images that cut across provincial boundaries, only the part falling into a particular province was considered to be in that province.

The area available for agriculture in the sampling frame covers irrigated and rain-fed land. The total area in the 18 provinces was 52,928 km^2, which is equivalent to 41% of all potential agricultural land in Afghanistan. Potential land refers to all land available for cultivation and also includes land that is currently fallow.

Cells containing less than 0.25 km^2 of potential agricultural land were excluded from the sampling frame in order to reduce the likelihood of choosing cells with very little arable land. In total, the exclusions represented less than 1% of the total potential agricultural land.

2.2 Sample size determination

The total number of images to be selected in the sampled provinces was determined in 2015 with the goal to increase accuracy of the estimates and to save cost when compared to previous years.

The accuracy of area estimates depends on the proportion of land covered by satellite imagery and even more so on the number of images than can be acquired. With opium poppy cultivation being concentrated in hot spots and thus unevenly distributed across the agricultural land, information from a large, contiguous piece of land has less value than geographically evenly distributed, smaller pieces information. Costs associated with satellite imagery depends mainly on the total area covered (and not on the number of images). By using 5 x 5 km instead of 10 x 10 km images, at same costs four times the number of images can be acquired. Further details on the sample size determination methodology can be found in O*pium Survey, December 2015*, page 42.

2.3 Sample size allocation

The available number n of images has been distributed to provinces h according to a so-called power allocation, which uses agricultural area as size measure. For provincial sample size n_h,

$$n_h = n \frac{X_h^q CV_h}{\sum_{h=1}^{H} X_h^q CV_h}$$

where CV_h is the coefficient of variation of area under poppy cultivation in province h and X_h land available for agriculture in province h. This approach ensures that sample size depends on both the variability of poppy and the size of the province measured by agricultural land. After an empirical assessment, the smoothing parameter q, $0 \leq q \leq 1$, was set to 0.2. In addition, a minimum of 20 samples per provinces was set, which took effect in Day-Kundi and Kunar.

Table 2 Sample size and agricultural land and sampling ratio, by province, 2024

Province	Total arable land (km2)	Frame	Effective sample size	Arable land in selected cells	% of arable land represented by selected cells
		# cells	# cells	(km2)	
Badakhshan	4,206	396	88	645	15%
Badghis	6,290	636	50	830	13%
Faryab	7,970	532	86	1,426	18%
Jawzjan	3,440	294	39	530	15%
Laghman	263	103	25	61	23%
Ghor	1,615	1144	83	114	7%
Day Kundi	672	406	20	29	4%
Farah	2,820	604	46	480	17%
Helmand	5,849	696	98	1,283	22%
Kandahar	3,472	695	80	763	22%
Kunar	293	124	24	50	17%
Nangarhar	1051	181	26	334	32%
Nimroz	1,083	213	28	237	22%
Balkh	4577	256	40	865	19%
Saripul	3557	379	56	600	17%
Takhar	3,402	456	60	1,006	30%
Uruzgan	903	277	30	89	10%
Zabul	1,465	541	29	150	10%
Total	**52,928**	**7,933**	**908**	**9,492**	**18%**

2.4 Sample design

In 2015, MCN/UNODC undertook an extensive simulation study which compared various sampling designs and estimation methods in order to determine the best (most accurate with a given number of samples) design for a certain situation.

Case studies were undertaken for Helmand and Kandahar province. The sampling designs considered have been used in the past by MCN/UNODC:

- simple random sampling,
- probability proportional to size sampling (PPS), using agricultural area as a size measure,
- stratified random sampling using compact geo-strata of equal size as strata,
- systematic random sampling.

Two estimation methods have been compared: a ratio estimator using agricultural area as auxiliary variable and the Horvitz-Thompson estimator.

The study concluded that for the two cases considered

- PPS performed best, and
- The ratio estimator is to be preferred for simple random sampling, systematic random sampling, and stratified random sampling. For PPS, it does not yield any improvements in accuracy.

The PPS builds on the correlation between the size measure and the variable of interest. In provinces where poppy and agricultural land are highly correlated, PPS is expected to perform best. In provinces, however, where poppy and agricultural land are only weakly correlated, PPS does not bring any advantages and might reduce accuracy. The drastic reduction in opium poppy cultivation since 2023 may have led in some cases to reduced accuracy of the estimation, as poppy has moved to more remote locations and away from the main agricultural areas.

In Badghis, Balkh, Farah, Faryab, Helmand, Kandahar, Nimroz, and Zabul, PPS was applied. In the remaining provinces, systematic random sampling was used, a sampling design that ensures an even geographical distribution of samples (see the "Opium poppy 2015 – Cultivation and production" for more details).

In more detail, in a PPS design without replacement a unit has a probability to be selected in the first draw of

$$pi = \frac{x_i}{\sum_{i=1}^{N} x_i}$$

where *x* is the size variable (agricultural land) in unit *i*, and *N* is the number of units that can be selected. The subsequent units have slightly modified inclusion probabilities. For drawing the samples and for calculating the inclusion probabilities the statistical software *R* (package *sampling*) was used.

Since agricultural area tends to be concentrated in one or more clusters in a province, PPS sampling without further stratification would lead to a concentration of samples in a few spots and possibly do not cover every district. Therefore, in all PPS provinces, the sample was stratified by district.

In the remaining provinces, a one-stage systematic random sampling approach was employed in which a sampling rule was applied that ensured good geographic coverage. Starting from a randomly chosen cell, every *k*th element from then onwards was chosen, where *k* is determined by the number of cells in the frame and the desired sample size (the actual sample size might differ slightly).

In *Nangarhar* province, the districts Dara-e-Nur, Kuzkunar, Kama, Behsud, Jalalabad and partially Surkhrod were excluded from the frame.

3. Area estimation in sampled provinces

The estimation of the extent of opium poppy cultivation is a ratio estimate[1] for each of the provinces, using potential agricultural land as an auxiliary variable. The national estimate was obtained by adding up the provincial estimates in what is known as a separate ratio estimate.

In provinces where systematic random sampling was applied, the area of opium poppy cultivation, Yk, within province *k,* is estimated as:

$$Y_k = X\frac{\sum_{i=1}^{n_k} y_i}{\sum_{i=1}^{n_k} x_i}$$

where n_k is the number of satellite image locations within the province; y_i is the area of poppy cultivation in image *i;* x_i is the area of land potentially available for poppy cultivation in image *I*, and *X* is the total potential land available for poppy cultivation in province *k*.

In PPS provinces, where units are selected with unequal inclusion probability, a slightly different ratio estimate was used that incorporates the inclusion probability (Horvitz-Thompson estimator).

3.1 Uncertainty

Since 2023 the drastic reduction in opium poppy cultivation led to a strongly skewed distribution of the observed opium poppy in the samples. Therefore, for all provinces where a sampling approach was applied, bootstrapping was used to estimate variance and the confidence intervals. Uncertainty at the provincial level was combined to assess uncertainty at the national level.

Area estimation in target provinces

In provinces where opium poppy cultivation has been highly concentrated in comparable little land, targeted satellite images were obtained to measure the areas involved. The total poppy area of a target province is

[1] The ratio estimator did not outperform the Horvitz Thompson estimator in the PPS provinces. The ratio estimator was applied in all provinces for reasons of consistency and to account for possible updates of the agricultural area in future years.

equal to the poppy area measured on the imagery without any further calculation. For a list of provinces for which the target approach was used see respective table.

In provinces where satellite images were targeted, the estimated area under opium poppy cultivation is not affected by sampling errors, although they may be affected by the omission of areas with very little cultivation. Area estimates of target provinces should therefore be considered as a minimum estimate.

District level estimation

Following the 'do no harm' principle, district-level estimates of poppy cultivation have not been provided since 2023 due to the decline in cultivation. Instead, a gridded system indicating relative poppy concentration has been adopted in 2024. This system involves a nationwide 500m² grid, where each cell's poppy production is measured and compared to the national average. Areas with above-average poppy concentration are thus highlighted, allowing for targeted analysis while maintaining a higher level of data sensitivity.

Accuracy assessment

Due to the difficult security situation in many parts of Afghanistan which prevented surveyors from carrying GPS it was not possible to conduct a systematic accuracy assessment. However, spot checks were conducted across the country to improve accuracy.

Estimation of the net cultivation area

In years where eradication took place and was verified by UNODC, the area figure has been the net harvestable opium poppy cultivation area. The effect of poppy eradication activities was taken into account based on data from the eradication verification survey, which provided exact GPS coordinates of all eradicated fields supplemented with additional information. The gross cultivation areas would be the net cultivation plus area eradicated.

In 2024, no in-house data was available to incorporate eradication efforts into the estimation of cultivated area. However, the DfA provided several datasets on eradication, which were analyzed. Although technical issues with the data prevented a full assessment of the impact of eradication, it was possible to determine the timing of these efforts. Findings showed that over 92% of reported eradication occurred before UNODC acquired satellite imagery, meaning nearly all eradication was accounted for in the poppy cultivation and opium production estimates.

4. Satellite image interpretation

Acquisition of satellite images

The acquisition of satellite images at the appropriate growth stage of the opium poppy is key to the successful identification of opium poppy fields on satellite images. In recent years, detailed information on the crop growth cycle of each district has been collected in the form of a phenological chart, which is useful for deciding on appropriate dates for satellite data acquisition. Images of the Southern, Eastern and Western regions are collected during March and April due to the early cultivation and maturity of crops in those regions. The crop growth cycle begins later as one goes northward and elevations are increased. Images of the North and North-eastern region are acquired during May, June and July.

The time window for satellite data acquisition is twenty days, depending on the scheduled passing of satellite and weather conditions. The time window for image acquisition aims to capture the full flowering stage. Images acquired in the middle of the prescribed time window facilitate optimum discrimination between opium poppy and other crops. In 2024, the usual acquisition windows were moved to adjust for the weather changes of the season.

The figure below illustrates the spectral characteristics (expressed in a Normalized Difference Vegetation Index; NDVI) of opium poppy and other crops between February and June. Wheat and opium poppy have the same growth cycle between March and June, as illustrated. The spectral differences between those two crops are more pronounced in February, which marks the beginning of the capsule stage of the crop in this example.

Figure 1 Spectral reflectance of opium poppy and other crops

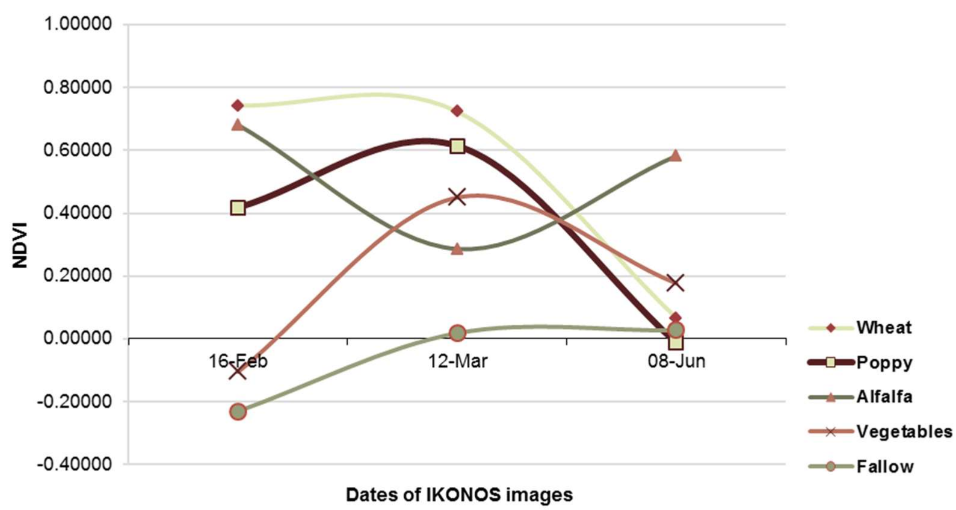
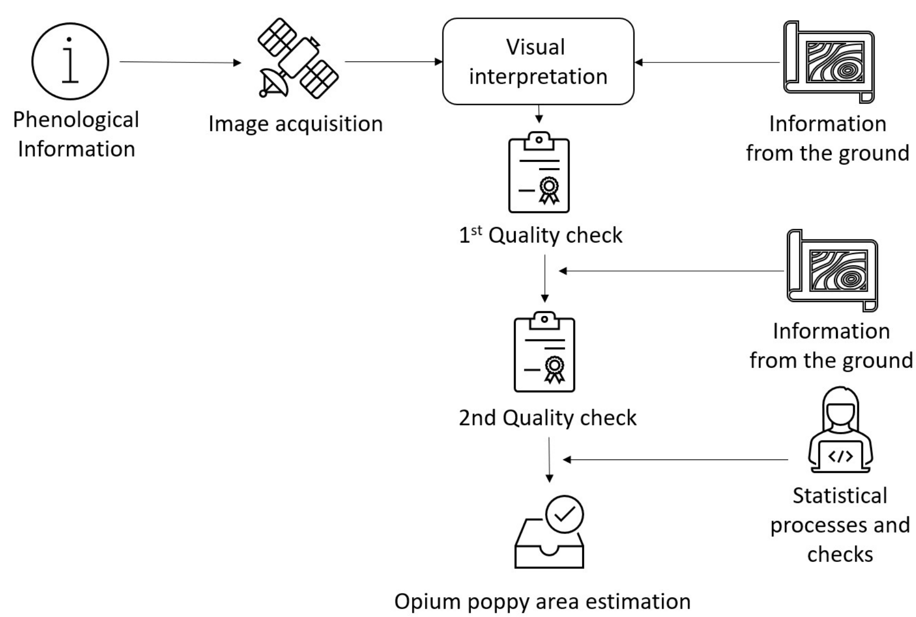


Figure 2 Image classification methodology for estimating opium poppy cultivation area



Interpretation of opium poppy cultivation from satellite images

Images were acquired during the flowering stage of the plant, when their spectral signature is clearest. This helps distinguishing opium poppy fields from other crops, such as wheat, which appears much redder when observed using false colour composite, or against other crops that may not have matured yet and will appear much paler.

Visual interpretation was used to delineate opium poppy fields by interpreting PLEIADES images covering a 5 km by 5 km area. Ortho-rectified PLEIADES images of 0.5 m resolution (PAN-sharpened) were used for this purpose. Opium poppy was identified using these high-resolution images.

Ground truth information was not possible to be collected this year due to the security and political situation on the country. However, images and reports from different sources were used to confirm of reject those cases where doubts existed about the presence of poppy. Poppy field boundaries were delineated by an on-screen digitization method.

Band combination for opium poppy identification

Two kinds of band combination were used to detect opium poppy. True-colour combination (blue, green, red) was employed with false-colour combination (infra-red, red, green) in all cases. Analysts used both combinations simultaneously to optimize discrimination between opium poppy and other crops.

Quality control

A quality control mechanism was applied to the image interpretation process, with each analyst's work being checked by at least two other experts, and in more complex cases three or even reviewers were involved in elucidating if certain fields were poppy or otherwise. All fields determined as likely to be under opium poppy cultivation (potential opium poppy fields) were delineated on the basis of the agreement of all interpreters involved. A final overarching quality control stage was introduced to make sure all fields were correctly included as part of the opium estimations.

5. Opium yield and production

In 2024, due to the continuing situation on the ground, it was not possible to collect opium yield data from field measurements. In the absence of the field survey, a method that adapted the the Global Agriculture Monitoring System (GLAM) system to the yield monitoring of opium was designed[2].

The methodology for forecasting poppy yields in Afghanistan in 2024 leverages a range of Earth Observation (EO) datasets, including NDVI, rainfall, temperature, soil moisture, and the Evaporative Stress Index (ESI). These datasets provide insights into vegetation health, climatic stress, and soil moisture, all critical for understanding poppy growth conditions. The data is processed to ensure compatibility with a poppy-specific crop mask and poppy crop calendars. Key climatic-impact drivers are created from this data, capturing agricultural conditions that affect yield. Correlation analysis is then performed to assess how these drivers align with historic yield data. The best-correlated indicators help build a yield forecasting model, specifically a CatBoost machine learning model, which accommodates complex, non-linear relationships between climate data and crop performance.

To refine the model, the CatBoost framework is calibrated through various yield scenarios, including assessments with individual data types to identify region-specific yield drivers. The final model's interpretability is enhanced with SHAP values, which identify the EO variables most impacting yield predictions in different regions. For instance, temperature emerges as a major factor in the southern regions, whereas soil moisture and ESI are more impactful in the northern areas. Confidence intervals for the predictions are generated using conformal prediction, providing reliable uncertainty estimates for the 2024 forecast. This combined approach enables a nuanced and regionally adaptive forecasting tool, designed to capture Afghanistan's varied poppy cultivation dynamics accurately

[2] https://glam.nasaharvest.org/info

Table 4 Yields (kg/ha) applied for each province for the 2024 opium season

Province	Yield
Badakhshan	37.7
Badghis	22.9
Baghlan	37.7
Balkh	34.8
Day Kundi	47.5
Farah	22.9
Faryab	34.8
Ghazni	24
Ghor	22.9

Province	Yield
Helmand	25.8
Hirat	22.9
Jawzjan	34.8
Kabul	47.5
Kandahar	25.8
Kapisa	47.5
Kunar	39.2
Kunduz	37.7
Laghman	39.2

Province	Yield
Logar	47.5
Nangarhar	39.2
Nimroz	25.8
Nuristan	39.2
Samangan	34.8
Saripul	34.8
Takhar	37.7
Uruzgan	25.8
Zabul	25.8

6. Crop change

In 2023, examining how crops change from year to year in Afghanistan focused on four provinces, Kandahar, Helmand, Nangahar, and Farah. These provinces collectively contributed 74% of the opium production in 2022, making them significant areas for observation. A cloud of approximately 9,700 sample points was randomly allocated within the opium poppy sample areas on agricultural lands. The number of points was calculated so it was representative sample of the agricultural practices of the studied provinces. In order to avoid coding the same field more than once, points were all located at least 50 m away from other points.

At each sample point, the identification of crops for both the current and preceding years was conducted using satellite imagery in line with the broader opium survey methodology. This standardized approach ensured a detailed analysis of the dynamics in crop types.

For 2024, given its current importance in poppy cultivation in Afghanistan, this exercise was expanded to also cover Badakhshan, with a further 2,587 sample points.

Table 5 Number of sample points per province

Province	Number of sample points
Badakhshan	2,587
Farah	2,134
Helmand	3,589
Kandahar	2,793
Nangahar	1,832