University of Southampton

Faculty of Engineering and Physical Sciences

Physics and Astronomy

Calibration of the CMS Electromagnetic Calorimeter with the EFlow Phi-Symmetry Method

by

Philip Backes

May 2025

Supervisor: Dr. Thomas Reis
Second Examiner: Prof. Alexander Belyaev

University of Southampton

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES
Physics and Astronomy

Master of Physics

by Philip Backes

The EFlow method of intercalibration of the CMS (Compact Muon Solenoid) Electromagnetic Calorimeter (ECAL) has already been developed for retroactive calculations after the data taking has been completed. It is a modification of the ϕ-symmetry method which makes use of the symmetry of the CMS detector and the initial conditions of the proton-proton collisions in the ϕ direction. These imply that the results of any experiment will (with a sufficiently large sample size) be symmetrical in the ϕ direction. This includes the transverse energy measured at each crystal in the ECAL, to provide corrections for the energy drift. In order to be able to execute these calculations efficiently during data taking and provide corrections of the regional drift in the response of the ECAL, in particular to the radiation ageing of the PN diodes used as references for the laser calibration, the EFlow calculations are integrated into the CMS automation framework. The framework's freedom, in how data is grouped is used to optimise the eflow intercalibration maximising the precision of the calibration.

Chapter 1

Background

1.1 The CMS Detector at the LHC

The Large Hadron Collider is a circular particle accelerator which accelerates charged particles to high energies up to 6.8 TeV. These particles are then collided inside detectors with the aim of producing collisions with energies up to 14 TeV. The detectors on the LHC are used to measure the properties of the particles produced by the collisions. From this the physical laws governing the production of the collision products, on the basis of the energy, trajectory, location and identity of the various products, can be tested.

The Compact Muon Solenoid (CMS) is one of the detectors on the LHC; it is cylindrical in shape around the beam line with the detector systems working outwards in concentric surfaces. As can be seen in figure 1.1, the systems, working out from the beam line, are the Silicon tracker followed by the Electromagnetic Calorimeter (ECAL) followed by the Hadronic Calorimeter (HCAL) then, (after the magnetic solenoid for which the CMS is named) the muon chambers.

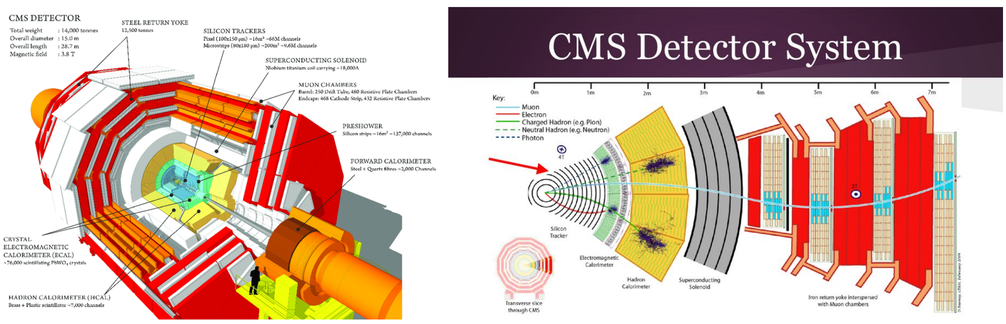

FIGURE 1.1: A pair of diagrams from [GC19] showing the structure of the CMS detector.

The Silicon Tracker is detailed in [DGV+04] measures the path of charged particles, so that estimates of the momentum, can be found. The location of a charged particle is found when it moves through silicon microstrip detectors. As particles travel through the tracker the pixels and layers of perpendicular microstrips produce electrical signals, the source of which locates the particle. The ECAL is made up of an array of 75,848 lead tungstate crystals; it measures the energy of electromagnetically interacting particles (the most common are photons and electrons) which are measured when they deposit their energy in the crystal, by scintillation, which the associated electronics measures. This is explained in further detail bellow. The hadronic calorimeter [HCA97] is made of repeating layers of dense absorber and tiles of plastic scintillator. When a hadronic particle hits a plate of absorber, an interaction can occur producing numerous secondary particles. These secondary particles flow through successive layers of absorber interacting to produce a shower of particles. The particles of this shower pass through the alternating layers of active scintillation material causing the emission of light. This light is amplified then measured by Silicon Photomultipliers to produce corresponding electrical signals indicating the location of the hadronic particle in the HCAL and their energy. A crystal hadronic calorimeter would provide a greater granularity than the current system hoverer the constrained volume of the CMS raises the cost too high.

The original purpose of the CMS detector was, to illuminate the origin of inertial mass, the most favoured candidate for which was the Higgs boson. Theoretical constraints put an upper bound on the Higgs mass and experimental results at the LEP collider (the LHC's predecessor at CERN) place a lower bound. Therefore it was known that if the rest of the standard model is correct up to the Plank scale the Higgs mass will be between 85 GeV and 420 GeV. The Higgs was found by the CMS and ATLAS experiments with a rest energy of 125 GeV. Now the CMS is conducting experiments looking for physics beyond the standard model (for example [SC22] so far unsuccessfully looking for supersymmetric particles) and looking at higher precision measurements of the standard model parameters, for example [CMS25] the top mass, which acts as a consistency test for the standard model, and higher precision tests of the Higgs.

1.2 The CMS ECAL

The CMS Electromagnetic Calorimeter (ECAL) [CMS97] measures the energy of electromagnetic particles produced by proton proton collisions within the CMS detector at the LHC. This is done using a cylindrical array of lead tungstate ($PbWO_4$) crystals. When an electromagnetically interacting particle enters into a crystal, its energy is transferred to the crystal and released

as light by scintillation; this then is collected by photodetectors, producing an electrical signal, with amplitude directly proportional to the energy of the particle. The geometry of these arrays mirrors that of the CMS as a whole where the Barrell region (EB) is the portion of the region which is cylindrical with the plane with 61200 crystals. The end-cap region (EE) of the ECAL is - as the name suggests - two circles with 7324 crystals each at the ends of the cylinder formed by the EB, with small holes at the centres for the LHC beam. The Angular direction perpendicular to the one of cylindrical symmetry (ϕ) is pseudorapidity given by $\eta = -ln(tan(\frac{\theta}{2}))$ (see [Won94]), where θ is the conventional angle in spherical coordinates. Pesudorapidity is used in high Energy Physics because differences in it are Lorentz invariant. The EB region covers the range $|\eta| < 1.479$ and the endcap covers the region $1.479 < |\eta| < 3$.

The crystals for each half-barrel are grouped into 18 "supermodules" each covering the same angle in ϕ. Each supermodule is made up of four modules where 500 crystals are in the first module and each of the other three modules contains 400 crystals. Each end-cap is organized in two semicircular structures, called "Dees". The scintillation light, produced in the crystals is detected by Avalanche Photodiodes (APDs) in the EB and by vacuum Phototriodes in the EE [cms24]. In front of each end-cap there are pre-shower detectors (ES) which are two planes of silicon sensors with a total of 3 radiation lengths of lead in front of each endcap, they cover the pesudorapidity region $1.653 < |\eta| < 2.6$.

The purpose of the ECAL is to provide precise measurement, including time and location, of the energies of electromagnetically interacting particles and their showers. One of the most significant examples of these is the observation of the decay products of the Higgs boson into di-photon or four lepton final states. In some of the proton-proton collisions occurring as a result of the bunch crossings at the centre of the CMS a Higgs boson will be produced. This then sometimes decays by the following:
$H \rightarrow \gamma\gamma$
or alternatively the Higgs boson may decay into a pair of neutral Z bosons each of which may decay into an electron positron pair
$H \rightarrow ZZ \rightarrow 2e^{+}2e^{-}$

The total energy of these decay products will be the same as the total energy of the Higgs boson, from which the rest energy can be found. So measurement of these decay products provides an indirect measurement of the mass of the Higgs boson. When the total energy distribution for the decay products is plotted, this appears as a Gaussian peak in the number of particle sets with that combined energy, centred at the Higgs boson mass, see [4le14].

This technique is applicable to any particle where you know and can measure the decay products.

1.2.1 ECAL Resolution

The precision of the CMS ECAL is described by its resolution which is given by [CMS97] pgs 304-309

$$\sigma^2 = a^2 E + \sigma_n^2 + c^2 E^2 \tag{1.1}$$

for energy measurements $25GeV < E < 500GeV$, additional terms can be added to this equation making it applicable at higher energies. a is the stochastic term which accounts for: fluctuations of the lateral containment of showers to the 5×5 crystal groups, which the reconstruction algorithm assumes that showers will be contained to; fluctuations in the energy absorbed by the preshower absorber as well as photostatic contribution which is inversely proportional to the number of primary photoelectrons released in the photodetector per GeV. The noise term σ_n describes the noise from the preamplifier, digitization (which is most significant for low energy measurements) and the pileup (which is insignificant at low luminosities). Pileup is the interactions between additional, mostly low-energy interactions between non-colliding protons in the same bunch crossing. The constant term c receives it's most significant contributions from the non-uniformity of the longitudinal light collection and per crystal intercalibration errors. The calibration method worked on for this thesis decreases the intercalibration errors and therefore the constant term leading to an improved energy resolution (smaller σ).

1.3 Crystal Darkening: the need for Continuous Calibration

During the operation of the LHC the $PbWO_4$ crystals are irradiated by both electromagnetic and hadronic radiation. These result in a darkening of the crystals, with the electromagnetic doing so by creating colour centres in the crystals, and the hadronic by shifting the transmission band energy. These darkening effects decrease the magnitude of the electric signal produced by the photodetectors and would erroneously change the measured energy over time, if left uncorrected. When the radiation ceases, during periods without collisions, the colour centres naturally heal by themselves; however, the transmission band does not naturally shift back, therefore there is a partial but incomplete recovery of the crystal transparency at times when there are no collisions. [AAK+98] study the mechanism for this in more detail.

The energy reconstruction requires a calibration to assign the correct value of the energy from the signal. The regional changes in this calibration can be seen in figure 1.2. These measurements are taken by directing laser light of known energy at the ECAL. The observed decrease in response (a result of the decrease in crystal transparency) occurs during data taking periods and there's also a partial recovery during periods where there's no radiation. The darkening and partial recovery of the crystals implies that, for accurate results, calibration must be repeated after each time interval when the previous calibration is no longer sufficiently accurate. Those periods are referred to as intervals of validity (IoVs), so an IoV is an interval of time for which a given calibration is still sufficiently accurate for physics analysis based on the calibration.

The requirements on calibration are made more complicated by the fact that (1) the radiation environment is not uniform along the detector and (2) the crystals are not all made identically. Therefore there is both a bulk change in the level of calibration required by the ECAL channels based on the region of the detector and local inhomogeneity among the individual crystals based on their different responses to the radiation. [Rog10] explain the calibration techniques used to combat this in LHC run 1 and [AC16] presents a study of these effects in a test beam.

1.4 Calibration Methods

1.4.1 Laser Calibration

As [cms24] explains, the particle energy measured at a given crystal may be calculated with the following equation:

$$\epsilon_{ch} = IC_{ch} \cdot LC_{ch}(t) \cdot G \cdot A_{ch} \tag{1.2}$$

where ϵ_{ch} is the energy of a single event measured by a crystal specified by the channel ch, IC_{ch} is the inter-calibration constant for that specific channel and $LC_{ch}(t)$ gives the time dependent laser correction. A_{ch} is the amplitude of the signal produced by the photodetector in ADC counts and the constant G converts this amplitude into physical energy units (conventionally MeV), from ADC counts, with approximate values of $40MeVADC^{-1}$ in the EB region and $60MeVADC^{-1}$ in the EE region of the detector. Calibration is the task of finding the IC_{ch} and $LC_{ch}(t)$ values so that an accurate measurement of the event energies can be made.

Therefore, the calibration process is conventionally broken up into two parts. In the first the laser calibration of the average crystal, in a given η region of the harness (see below), is found in order to account for the changes in the ECAL sensitivity over time. And in the second the per crystal inter-calibration. The first makes use of the laser correction data as in Figure 1.2 with these calibration corrections being given by:

$$LC_{ch}(t) = (\frac{R_{ch}(t=0)}{R_{ch}(t)})^{\alpha} \tag{1.3}$$

where $R_{ch}(t)$ is the response to laser light relative to the initial luminosity response at the start of the data taking period. These are shown in figure 1.2 and α is determined by the different laser and crystal emission light paths. Typically α has values around 1.5 in the barrel region (EB) and 1 in the end-cap region (EE).

$R_{ch}(t)$ is measured by a light-monitoring system using lasers to measure each crystal's transparency and the photodetector's response. A laser of wavelength of 447 nm (blue light) is used to measure and correct for changes in crystal transparency and photodetector response. Additional lasers with different wavelengths provide additional crystal transparency measurements that are used to check the accuracy of the blue laser transparency measurements and its applicability across the energy spectrum. The laser is injected into each of the crystals through optical fibres, with a final splitter breaking up the light into 200 separate fibres. Each fibre is directed at a single crystal, and two additional fibres injecting light into a pair of PN diodes which are used to measure the amplitude of the laser light pulse. These 202 fibres are called a "harness". The transparency is inferred from the ratio of the light measured by the photodetectors and the PN diodes, therefore the transparency estimate is not distorted by variations in the laser amplitudes between pulses as these variations apply equally to the photodetector channels as to the PN diodes.

1.4.2 PN Corrections

However the PN diodes also decrease in responsiveness to laser light as a result of LHC irradiation, but at a different rate measured by [GAB$^+$99]. Therefore an additional calibration is necessary to correct the resulting energy drift. For LHC Run 2 and Run 3 this was done with E/p corrections. These act at the granularity of the harness by fitting the distribution of energy of electrons from W and Z decays to a template to find the relative energy scale averaged over the harness region. The templates used vary according to the location along $|\eta|$ of harness region with regions at higher $|\eta|$ having broader distributions.

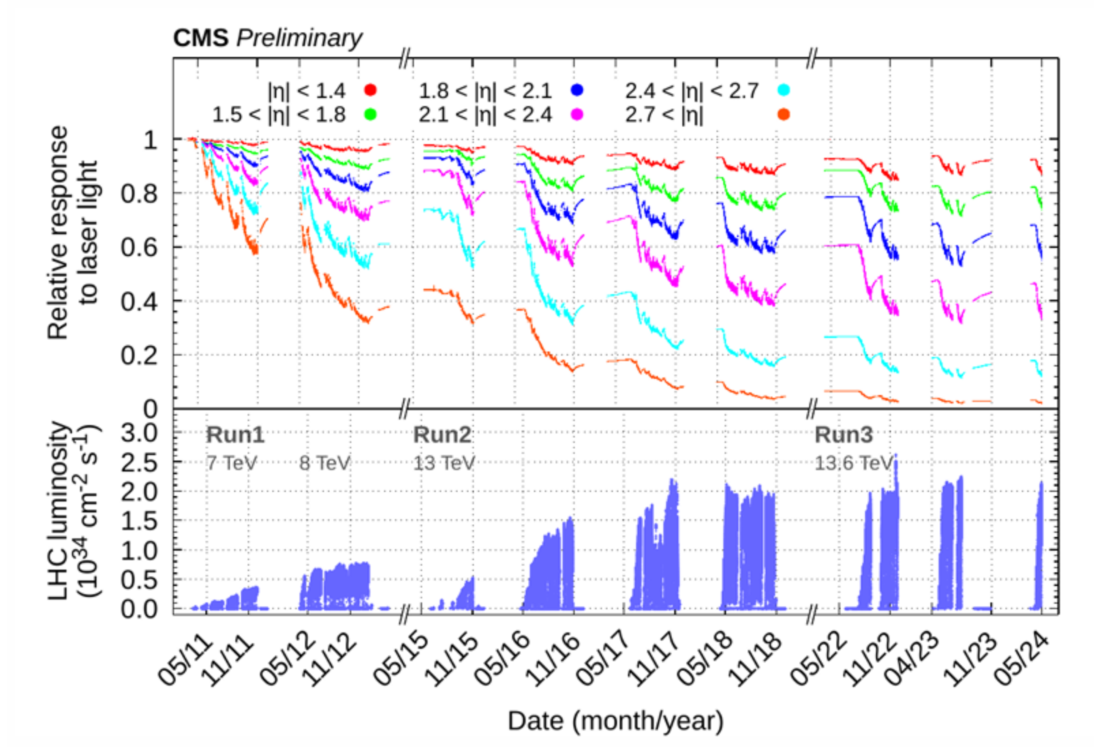

FIGURE 1.2: From [las19]. Shows the decrease in electrical response to scintillating laser light on the ECAL of each region of the the ECAL system from 2011 to 2024, this is a result of crystal darkening so relative response decreases because more light is being absorbed by the crystal. .

The EFlow calibration is an alternative to the E/p corrections with the advantage that it has the capacity to provide per crystal corrections to the drift in energy scale resulting form PN ageing compared to E/p which works at the granularity of the harness region.

1.4.3 Inter-calibrations

Methods used to derive the inter-calibrations (the implementation of which is explained in [cms24]) include the: π^0, E/p, Zee and ϕ-symmetry. These four are independent since they make use of different data selections from CMS targeting different phenomena. The π^0 method uses data from events producing two photons with energy in a window of invariant mass compatible with a measurement of the π^0 meson. From the energies of these decay photons (similarly to the Higgs decays above) the mass of the π^0 is measured. The inter-calibration constant is found by equalising the inter-calibrations so the measured mass is independent of where in the ECAL the photons are measured. The E/p method takes data for high energy electrons and compares the energy measured in the ECAL to the transverse momentum measured in the tracker, as the electrons move through the tracker. From these measurements the value of the electrons kinetic (and therefore total) energy may be inferred which should be equal to the energy measured by the ECAL (except for loss to Bremsstrahlung which is accounted for by the

e γ reconstruction algorithms [C+21]). The Zee method makes use of the electron invariant mass for electrons produced by a Z boson decay, finding a calibration in a similar way as is done from the π^0 decay products. The ϕ-symmetry inter-calibration uses events produced in low energy collisions, which are expected to, over a large data sample, have a uniform total energy distribution in the ϕ direction. Therefore, for a sufficiently large sample the average transverse energy deposited in a given crystal is the same as the average transverse energy deposited in all the crystals sharing the geometrical ϕ-symmetry. From this each crystal energy is rescaled to be in line with the average energy of its own η ring.

1.5 High Luminosity LHC (HL-LHC)

To provide larger datasets for searches for new physics and increased precision measurements of the standard model, in particular Higgs Boson couplings, the LHC (and therefore CMS) will undergo a series of upgrades. The central focus of these improvements will be an increase in the collision rate that delivers 250 fb^{-1} per year with a peak luminosity of $5 \times 10^{34} cm^{-2} s^{-1}$ compared to the 100 fb^{-1} with a peak luminosity of $2 \times 10^{34} cm^{-2} s^{-1}$ delivered for each of the years in run 3 after the phase 1 upgrades.

There are many hardware upgrades planned for CMS, including the ECAL ([Cc+24]), which will provide the required capacity for the increased instantaneous luminosity. One example of this is the increase in the time resolution of the ECAL required to distinguish measurements originating from collisions which are closer together in time than the current electronics would be able to distinguish. However, the increased rate of collisions in the detector will increase the rate of crystal darkening in the ECAL discussed in section 1.3. This will worsen the precision of the ECAL measurements unless the calibration is improved to compensate. One of the improvements to the ECAL calibration is the automation of the EFlow intercalibration detailed here (see [CMS17] pgs 106-107).

1.6 The ECAL Automation System

Presented in more detail in [Pig23], the CMS automation framework was developed during the second LHC long shutdown in anticipation of the increased calibration requirements as a result of the ageing of the ECAL and the increased rate of crystal darkening which will occur with the introduction of the High-Luminosity LHC (HL-LHC) program (see [CMS17] pg 8), for the entirety of which CMS will run. This will lead to an increase in the frequency with which the ECAL needs to be calibrated.

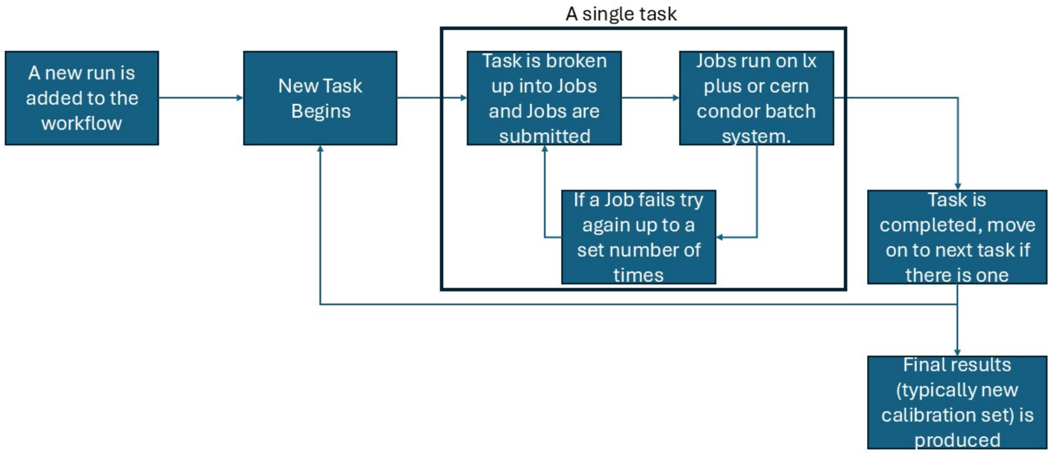

FIGURE 1.3: A flow chart showing the automation workflow in the simplest case where each task is performed with one job per run. Jobs within tasks are retried if they fail up to a specified number of times. Each task either works directly from CMS raw data or from data produced by a previous task's analysis, in this case the task producing the prerequisite data is strictly a prerequisite in the automation framework.

For the time spent on calibration to be used efficiently, the minimisation of human time dedicated to the execution of calibration calculations is required. Given the setup at the time, the next step of this was to automate the production of new calibrations during data taking, it also has the capacity to re-calculate calibrations after the completion of data taking for the improvement of legacy datasets, however the distinction between these is not important at this stage for the EFlow calibration.

The automation system makes use of: Jenkins (see [jen25]) for scheduling execution of the workflow tasks in the procedure; Grafana (see [gra25]) to display the status of the automated workflows and an InfluxDB (see [inf25]) database records the state of tasks in the various workflows and the locations of any output files of the task. Each workflow's process code is stored in the automation CERN GitLab repository. Typically a workflow task refers to calibration code in external repositories outside of the automation repository. The process code executes the calibration script and updates the database, interacting with it by the automation control module [eca25b]. Grafana displays the status of the database.

The CMS data acquisition system groups data from chronologically adjacent events into coherent units called runs. During an LHC fill several runs can be taken where the condition may differ between runs. A set of automation tasks and their associated files are contained within a campaign. Here each

task has a status [new, processing, done, failed, reprocessing, skipped] for each run. Each task will have one or more Jobs with the possible status being [Done, Failed, Running, Idle] and the automation framework indexing jobs by run number.

The sequence of the events in task completion through a generic workflow is shown in figure 1.3. In data-taking periods, new runs are injected automatically when the CMS Data Acquisition System marks the run as completed. Jenkins runs once an hour to check for completed runs then injects them into the InfluxDB database, this sets the workflow to be executed for that run. The automation framework also has the capacity to have runs manually injected into campaigns for processing. This can be used to re-analyse older data for recalibration re-reconstructions and similar analysis. In the case of recalibration, the relevant data is injected manually by users with the ECAL automation control python package [eca25b]. All the tasks in a given workflow are executed by Jenkins. If the status of a run in the workflow is "new" or "reprocessing" the task handler submits the jobs for that task then marks task's status as "processing". Tasks can also accumulate multiple runs if the data is grouped by: fill, a minimum integrated luminosity or a minimum number of events.

The database is connected to an interface to access the CMS data which is stored on the WLGC (Worldwide LHC Computing Grid [WLG25]). Output files from the automation tasks are recorded by the InfluxDB database for easy recalling of data by subsequent tasks with the python interface. Each job's output is typically a file or set of files stored on a shared filesystem within the CERN file storage network, with the location being saved in the InfluxDB. The database may also have values stored directly.

Chapter 2

Inter-Calibration Constant Calculation

2.1 The ϕ-symmetry Method

The cylindrical symmetry of the CMS detector and of the average initial state of the proton collisions produced by the LHC, gives rise to a uniformity of physics in the ϕ direction. This is used in the ϕ-symmetry calibration. This method specifically uses the expectation that the average energy deposited in a given crystal will be independent of that crystal's location in ϕ and assumes at least a rough calibration of the detector so estimates of the particle energies must already exist. This calibration prior to the ϕ-symmetry can be done only using the laser calibrations with a simplification of equation 1.2.

$$\epsilon_{ch} = LC_{ch}(t) \cdot G \cdot A_{ch} \tag{2.1}$$

The ϕ-symmetry method uses this to find the inter-calibration constants according to the following equation (given in [cms24] Appendix B.1):

$$IC_{ch} = \frac{\sum_j E_T^j}{N E_T^{ch}} \kappa \tag{2.2}$$

where the j sum is over the N crystals in the chosen η ring, κ is a constant accounting for the data filtering, and E_T^j is the measured transverse energy deposited in crystal j. This is then applied to every η ring in turn. The average may only take the values from the same η ring because the geometry of the CMS ECAL is not η symmetric about the collision point, and the physics of the collisions is not η symmetric, therefore, the distribution of energy deposition is not uniform in this direction.

The inter-calibration constants improve the energy estimate of the ECAL by the following:

$$\epsilon_T'^j = IC_j \cdot \epsilon_T^j \tag{2.3}$$

where $\epsilon_T'^j$ is an improved estimate of the single particle transverse energy ϵ_T^j measured at channel j prior to the ϕ-symmetry calibration. The ϕ symmetry calibration is typically used for monitoring of the ECAL calibration rather than performing it since is has a precision of around 2-5% whereas the π^0 and Zee methods have precision around 0.5 - 1.5 %.

2.2 Mathematics of the EFlow Method

The modification made to ϕ-symmetry to get the EFlow method, developed in [C$^+$22] chapter 3, is that the η distribution in particle frequency is found from data for the electrons produced by Z and W boson decays. This gives the energy distribution summed over events since the energy distribution in η, for the $Z \to e^+e^-$ and $W \to e\nu_e$ decay products [AB21], is also found from data. This is then used to weight the average of the energies deposited in the crystals. This allows the inter-calibration constants to be informed by the energy over the whole barrel region. This gives the EFlow method an advantage over the ϕ-symmetry method that shorter intervals are sufficient to get a large enough sample for there to be no non negligible random fluctuations in total particle energy for a given crystal. The modification due to the weighting is shown in the equation:

$$IC_{ch} = \frac{\sum_{j \in EB} \omega_j(i\eta) E_T^j}{N E_T^{ch}} \kappa \tag{2.4}$$

with the $j \in EB$ notation referring to the sum including all the crystals in the barrel region, and $\omega_j(\eta) E_T^j$ being the weight which is directly proportional to the number of electrons from Z boson decays collected by crystals in the corresponding η ring, calculated by:

$$\omega(i\eta) = \frac{n(i\eta)}{E(i\eta)} \cdot \frac{\sum_{i\eta} E(i\eta)}{\sum_{i\eta} n(i\eta)} \tag{2.5}$$

Where $E(\eta)$ is the total energy from events selected for the ϕ symmetry calibration on the η ring and $n(i\eta)$ is the number of electrons from W/Z decays landing on that $i\eta$ ring. The multiplication by the distribution:

$$\frac{n(i\eta)}{\mu(n(i\eta))} \tag{2.6}$$

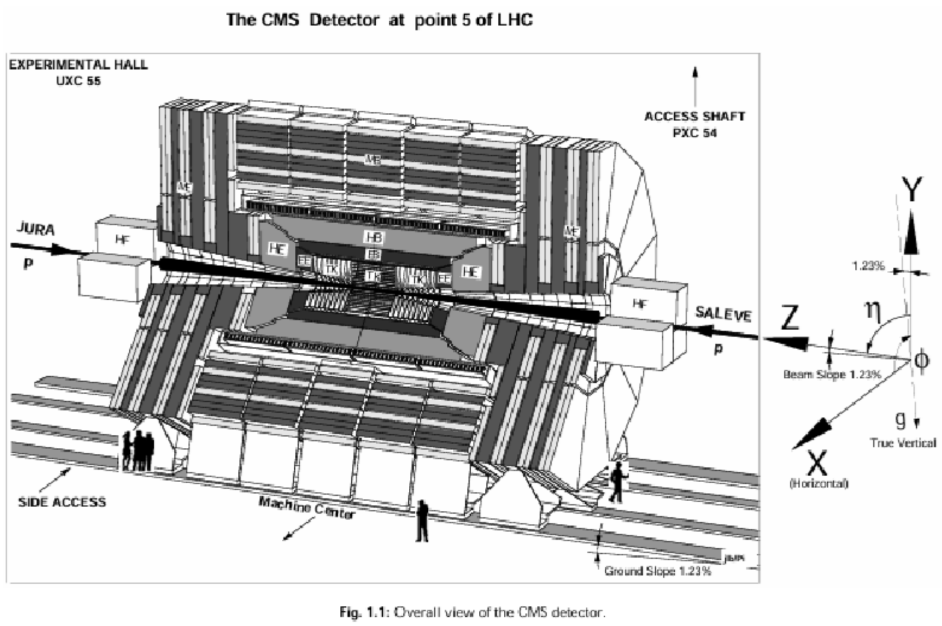

FIGURE 2.1: An overall view of the CMS detector from [CMS97]. The ϕ and η axis referred to are the ones shown on the right. The ϕ-symmetry of the detector around the z-axis, which is the beam line, can be clearly seen.

provides a normalised uniform distribution correcting for the non-uniformity of the crystal channels distribution along η. Then dividing by the distribution:

$$\frac{E(i\eta)}{\mu(E(i\eta))} \tag{2.7}$$

scales that flat distribution in η along the $i\eta$ axis to the data selection made for the transverse energy measurements used in equation 2.4, correcting for the non uniformity of Physics in η. The calibration constants found here correct the previous energy estimate the same way as in the ϕ-symmetry method, according to equation 2.3.

For this calibration method to be useful the IC_{ch}s must be recalculated periodically to keep up with the changes in the relative crystal transparency. The interval of validity (IoV) is the period of time for which a given intercalibration is accurate. This will include but extends beyond the data period from which the set of inter-calibration constants are calculated. When the intercalibration is found during data-taking it's typically applied for runs after the data period because of the time it takes for the calculations to be done, this only results in inaccuracies if these runs are outside the IoV. The data period contains all the events from the minimum number of runs necessary for the data period to contain at least a minimum integrated luminosity, which is the determiner of the rate of crystal darkening, so within each data

period a similar amount of crystal darkening occurs. The ICs are calculated from equation 2.4 independently for each data period.

This calibration is only applicable to the EB region of the ECAL because of the preshower detectors in front of the end-caps. The purpose of the preshower is to distinguish γ and π^0 particles. It is not symmetric in ϕ so a ϕ symmetry cannot be assumed. Until the endcap hardware is replaced this is not a significant concern, since the response of the endcap crystal channels is very small (see the high η region in figure 1.3) giving these measurements, even once calibrated, poor precision.

2.3 Data Filtering and Reconstruction

Since CMS collects more data than it can save, data useful for calibration and analysis is filtered. The first stage of this is a hardware filter (L1T) that is triggered upon specified conditions and the second is a software trigger (HLT) which looks for more specific physics information.

The data selected for the EFlow analysis are the same as for the ϕ-symmetry method. The aim is to ensure any statistical fluctuations away from the symmetry in ϕ are as small as possible. Therefore, the products of lower-energy interactions are chosen to prevent any rare high energy products from significantly skewing the data. In addition to this, from each event, only signal from ECAL (for example nothing from the tracker) above an energy threshold is saved in order to reduce the data storage space required per event, allowing more events to be stored and ensuring the stored measurements correspond to true particle events.

The events are select random colision events at the L1T. The selected data is then passed to the high level trigger (HLT). A random triggering at the L1T from the bunch crossing events is used ensuring there is no bias in the direction of ECAL hits introduced at the L1T. At the HLT events are selected for at least one crystal signal above 7 times the noise expectation. Subsequent to this an offline selection is performed where the minimum energy threshold is 10$\times$ the RMS of the noise and any events with energy more than 1 GeV above this are excluded.

2.4 The κ Term

The mentioned offline energy selection window is why the κ correction factors in equations 2.2 and 2.4 are necessary. Any miscalibration prior to the

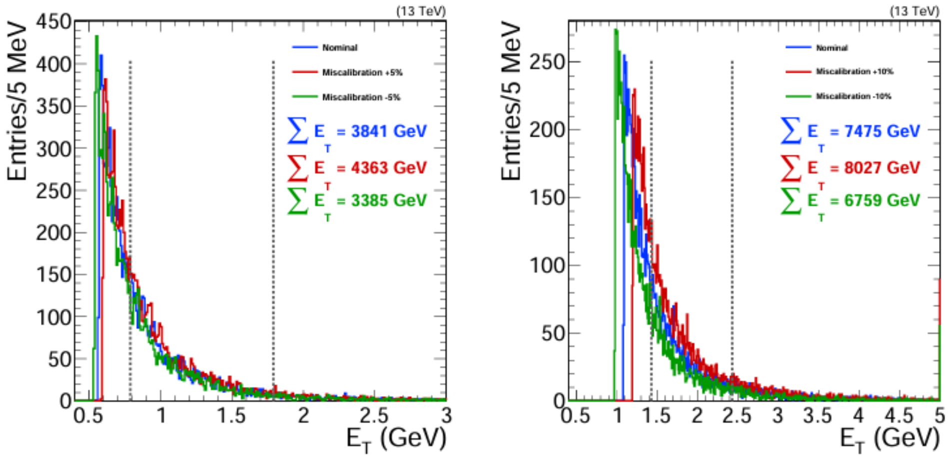

FIGURE 2.2: Deposited energy spectrum of events selected with applied miscalibrations. In the offline filter the same limits on the energy are used for selection (vertical dotted lines). The left figure shows the spectra for a ±5% miscalibration for the EB and the right shows spectra resulting from a ±10% miscalibration from the EE data. The sums of transverse energies given are within the filter range shown. [B+20] pg 65

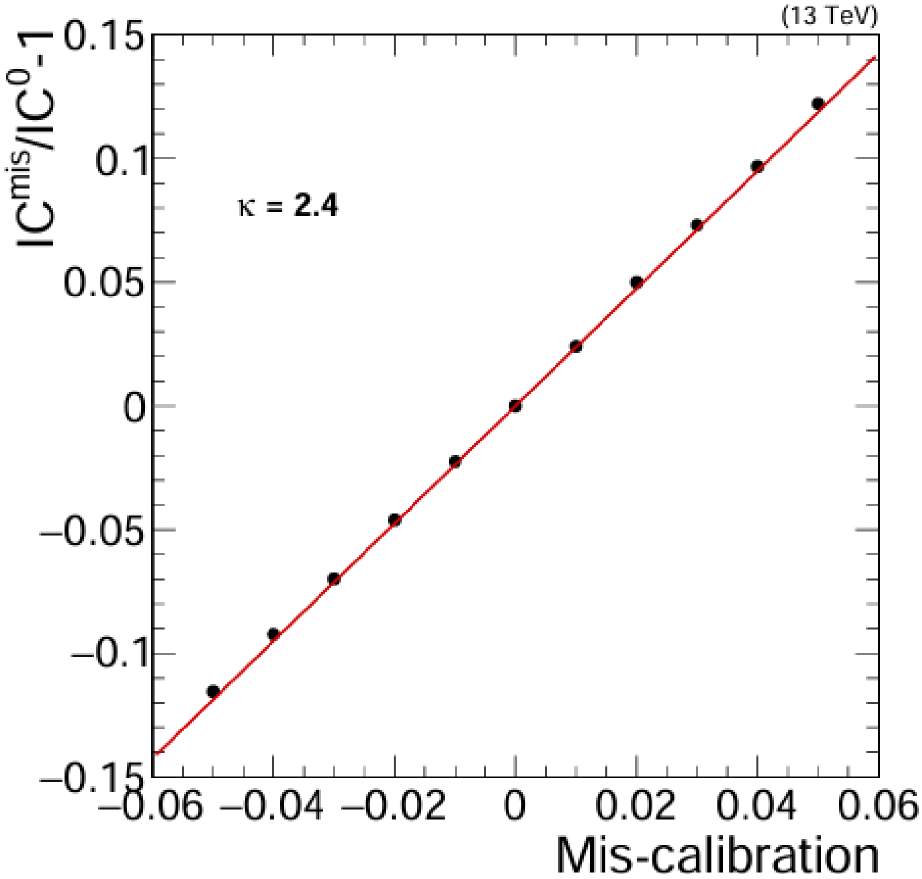

FIGURE 2.3: After the results shown of which an example is given in figure 2.2 the inter-calibration constants that would be implied are found. This is IC^{mis}. $\frac{IC^{mis}}{IC^{0}} - 1$ measures the proportional change in the calibration constant as a result of the miscalibration. [B+20] pg 66.

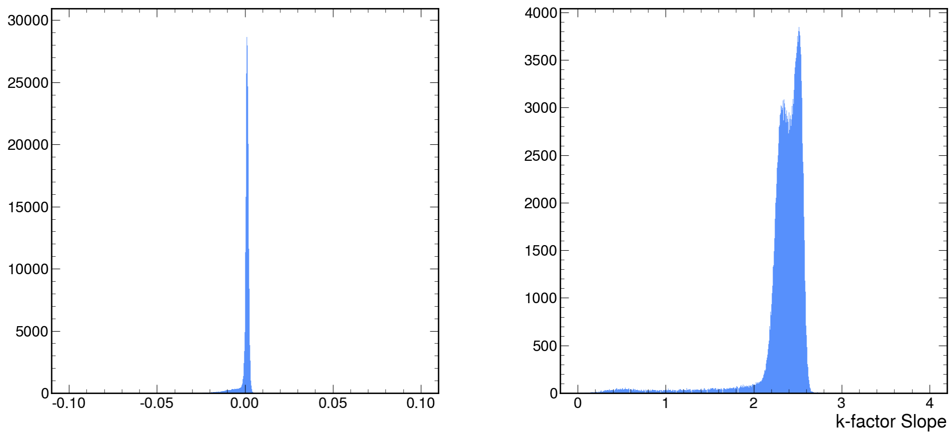

FIGURE 2.4: Frequency Histograms for the values found for the κ term (the slope of the graph figure 2.3) on the right and the intercept of the linear fit. The values are calculated once for each $PbWO_4$ crystal in the calorimeter.

EFlow inter-calibration produces an error on the estimated average energy at a crystal because it can affect whether or not an event is accepted by the selection window. This is corrected by the kappa factor. It is found by multiplying by a set of known miscalibrations to the energy data and re-applying the same energy window then finding what the inter-calibration constants would be. κ is the rate of change of the proportion of inter-calibration constant change with respect to the miscalibration. This rate of change is found from a linear fit finding the rate of change of the IC values with the intentional miscalibration. [B+20] illustrates this calculation with the figures 2.2 and 2.3.

The κ constant is dimensionless since both the proportional inter-calibration constant change and the miscalibration are dimensionless. Typical values for κ are around 0 - 3, and are shown in the right pane of figure 2.4. As would be expected the κ intercept is almost always 0 because when there is no added miscalibration the calculated inter-calibration constant is the same as the original inter-calibration constant. This simplifies the necessary accounting for the miscalibration since it can be assumed that κ_0 is exactly 0.

2.5 EFlow Automation

2.5.1 Benefits of Calibration Automation

If the EFlow calculations are done together with the initial calibration during and immediately after data taking, the need for repeated re-reconstruction is decreased; the new work presented here is exactly that, integrating the calculation software produced by [C+22] into the CMS ECAL calibration

automation framework [Pig23] so that the analysis is done automatically and the results can be used to improve the accuracy of the reconstruction of the physics.

The limitation in the ϕ symmetry and EFlow implementation of [B+20] and [C+22] is that it requires manual execution for the weights and eflow tasks to produce the inter-calibration constants; if the tasks are brought into the automation framework this requirement on human resources is removed. It also improves the time efficiency of the progress of calculations since the pointing of calculations to required data files is all automated so no human has to find them for calculation and standardises the exact way in which the calculations are performed improving reproducibility.

2.5.2 Workflow Organisation

The calculations for EFlow are done in two stages following [C+22]. First, the "weights" are calculated from the data outlined in section 2.2 and then the κ values followed by the inter-calibration constants are found. This break is made because the weights of the ECAL crystals are a feature of the detector implying that once a set of weights has been found from a dataset this is valid for a very large number of runs, with the possible exception of very slow changes arising from changing detector conditions such as new ECAL channels photoelectronics dying or ones which had been masked, and therefore returned no data, being uncovered. Following this break up of the workflow in prior work the automated workflow is similarly split in 2. The task where the weight values are calculated is called eflow-weights and the ICs calculation task is refereed to as eflow-eflow. eflow-cml collects the results for past runs of eflow-eflow normalises the calibration constants to the first set available and outputs the data in a data form from which the calibration can be implemented. Although not strictly necessary in terms of the relevant physics, the simplest way to construct the automation workflow is so that each EFlow stage requires as a prerequisite for running, and uses for its calculations, the weights values from the data for the same set of runs. The weights calculation both doesn't require as much data for statistical significance and has less intrinsic change over time than the eflow-eflow calculations.

The frequency with which the eflow-weights task is ran has a very small effect on the runtime since the eflow-weights calculations are around 80 times faster with greater disparity when the data period has a higher integrated luminosity. In addition to this, the storage space required is around 300 times smaller than for the eflow-eflow and eflow-cml. The benefit is the simplicity and efficient progress of the workflow with this structure compared

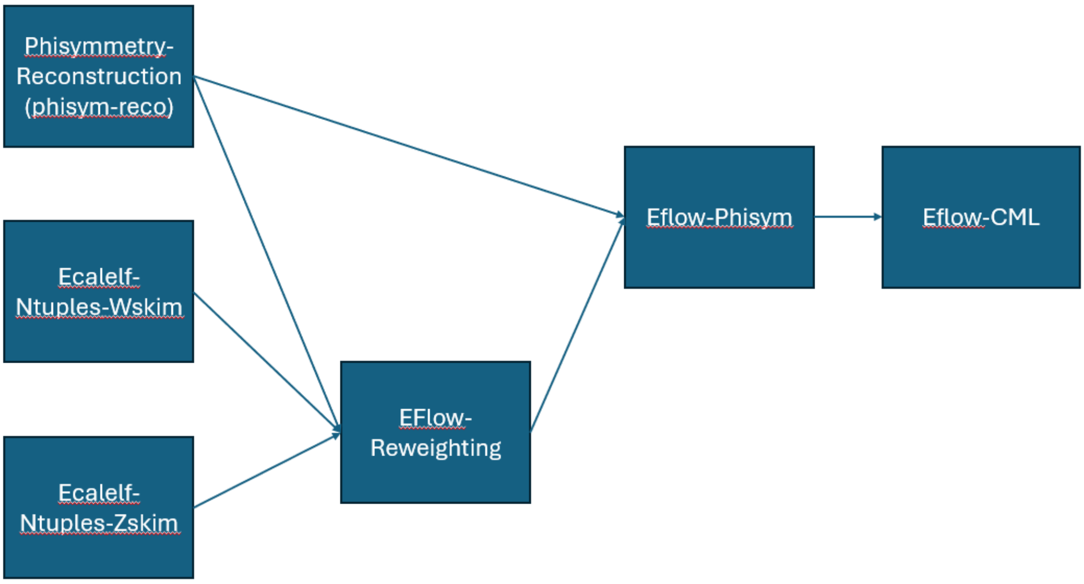

FIGURE 2.5: The task dependence order of the EFlow tasks as implemented in the automation framework. Tasks on the far left depend only on CMS reconstructed data; an arrow $A \rightarrow B$ indicates that A is a prerequesite for B. In all the cases here this is because the dependent task makes use of the data files produced by the preceding task for its calculations.

to if multiple different eflow-eflow data periods pointed back to one weights calculation.

There are seven tasks in the automation relevant to the EFlow calibration. Their dependence is illustrated in the figure 2.5. The phisym-reco task takes reconstructed CMS data and performs the offline data selection described in 2.3. From these results this task also finds the total energies deposited in the crystals over the given interval (no more than one run) and applies the miscalibrations to these results which will be used in finding the κ values, though they're found later. The ecalelf-ntuple-wskim and zskim tasks also work direct from the reconstructed CMS data with events selected for W or Z bosons. The analysis useful for the reweighting is the counting of electrons (and positrons) along the η axis with the specific case of interest being those electrons emitted by Z or W boson decays. These are used in EFlow-Reweighting to find the weight value, as explained above. The eflow-weights task produces the weights using the eflow-eflow task as in equation 2.4. The eflow-eflow task completes this by taking the weights from eflow-Weights in addition to the total crystal energies including the intentionally miscalibrated ones to find the kappa values and from them the inter-calibration constants. The eflow-cml task takes the results of all previous eflow-eflow tasks in the campaign and plots them together normalising them so that the IC value for the first data period under consideration is 1 for every crystal, and saves the data in a form useful for calibration. The patterns in the change of the IC values are shown which elucidates the changing properties of the detector channels.

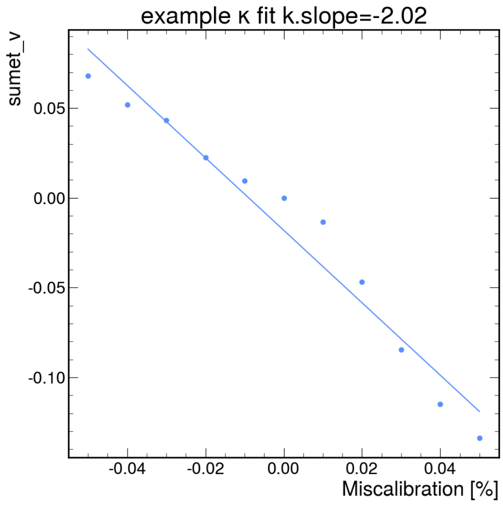

FIGURE 2.6: A linear fit to find the rate of change of the average crystal energy with the miscalibration.

2.6 Data Cleaning

There are a small number of cases where the calculation workflow returns nonsensical or invalid values. In many but not all cases this is because the ECAL channel is dead or masked but in some it is a result of unusual energy distributions on that crystal.

For masked channels (which are indistinguishable from dead ones) it is desirable to still have a calibration constant so that if it is uncovered energy measurements can be made, the accuracy of which will be improved upon by subsequent calibrations.

The most common, real but clearly invalid, results returned are negative intercalibration values. These come when the energy distribution of the hits on a crystal, like is shown in figure 2.2, is such that the average energy in the selection window changes in the opposite direction to the energy miscalibration. This gives a κ fit looking like figure 2.6. Through equation 2.4 this implies a negative intercalibration constant which is inconsistent with the design of the ECAL because the ECAL does not return negative energy values.

To give a viable IC value to these channels the average value is found for all the good crystals in their η ring and the channel is assigned the IC value of it's own η ring. This is an improvement over giving them a value of 1 since it responds to the η dependence of the intercalibration. The importance of this can be seen from figure 1.2 which shows the η dependence of the crystal darkening rate.

2.7 Optimisation

2.7.1 Length of the weights data grouping?

It still needs to be determined how best to organise the automation workflow so that the weights values are calculated sufficiently accurately for the subsequent calibration. The values should be very similar for all data periods so the only concerns are the: (1) simplicity of the automation workflow to decrease the need for human intervention (2) calculation of accurate weight values and (3) minimisation of the computational cost in storage and calculation time. The computational resources used by the eflow-weights task are very small compared to the eflow-eflow task, so only the first two considerations are significant.

It is more representative of both the statistical significance of the available data and the changes in the true ECAL calibration to measure the length of the data group in integrated luminosity than number of runs or fills, because the integrated luminosity directly indicates the amount of available data from the $Z \rightarrow e^+e^-$ decays. Run and fill groupings of data would vary in size because the amount of data in a given run and the number of runs in a fill vary significantly. The mean squared difference between two weight data sets is given by the equation:

$$E_{rr} = \frac{1}{N} \sum_{i\eta=-\frac{N}{2}}^{\frac{N}{2}} (\omega_1(i\eta) - \omega_2(i\eta))^2 \tag{2.8}$$

where the sum runs over the N different indices for the η values pointing to each ring of ECAL crystals. $\omega_1(i\eta)$ denotes the weight value for the crystal ring at $i\eta$ calculated from the first data set and the equivalent for $\omega_2(i\eta)$ with the second data set.

The weights calculation making use of the maximum amount of data is assumed to be the most accurate - treating this as the true values attempting to be reproduced allows the difference between the less accurate sets of values and this one to be treated like an error - so the mean squared error of each set of the weights values is found from the weights found from all of the different sets of input runs. Any systematic uncertainties can not be measured by this technique however this is not of concern here since changing the data group size would not change the systematic error anyway. The decision based on these findings is how to group the runs in the weights calculation so as to (along side the other aims given above) minimise the error resulting from an insufficient sample size of the data, so it is not attempted to find the systematic errors here.

2.7.2 Length of the data period for the eflow-eflow calculation?

In the final calculation of the inter-calibration constants from equation 2.4 the only degree of freedom available to be changed is the length of the data period. A longer data period increases the accuracy of the assumption of symmetry in the ϕ direction however a shorter data period increases the internal consistency of the data, over which the calculations assume there is no change in the detector conditions, and therefore less change in the true calibration of each crystal channel.

Since both the amount of data available for the averages in equation 2.4 and the extent of crystal darkening are determined by the integrated luminosity in the CMS detector over the time interval under consideration it's more consistent to define the interval of validity in terms of integrated luminosity rather than number of runs or fils. This is equivalent to number of hits as done by [C$^+$22]. The simplest case of the execution of the automation framework is when it's done by run or fill, the use of this would increase the efficiency of the workflow as it would match the phisym-reco and ecalelf-ntuples-(w/z)skim tasks. However for the eflow calculations more than one run of data is necessary for accurate results so this is not a viable option for the eflow automation.

The data grouping is done in [eca25b] by starting from an initial run adding consecutive runs to the group until the integrated luminosity is greater than or equal to the target luminosity, then a new group is started and the process is repeated. This leaves the remaining work of finding the optimal integrated luminosity to determine the lengths of the data periods over which the intercalibration constants are found.

Unlike in the weights case, there isn't a result which, is known to be the most accurate one by which to measure other possibilities. However we can get an idea of the optimal range of the data group length by plotting the results for the same crystals, over the same runs on top of each other. A sufficiently large data grouping can be seen when the results follow the systematic change, so the IC values move together with minimal noise for different but similar choice of data group length. A sufficiently small data grouping can be seen when the gaps between consecutive IC values are not significantly larger than the precision requirement of the calibration.

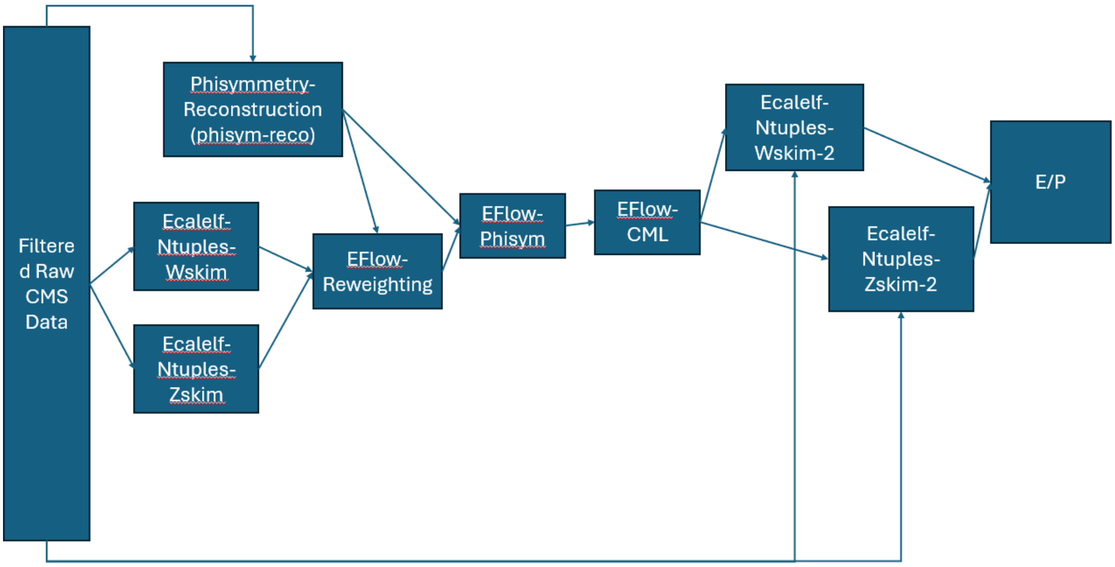

FIGURE 2.7: The automation framework dependencies for implementing the E/P calibration making use of improved ntuple data after the EFlow inter-calibration.

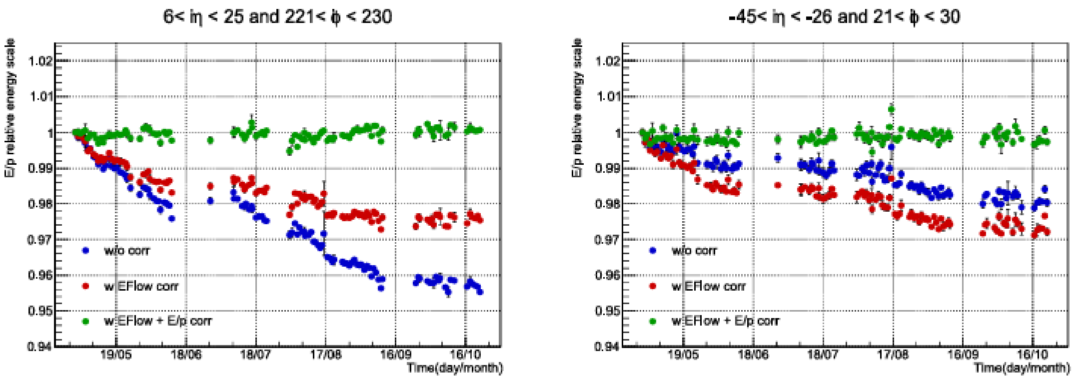

FIGURE 2.8: Comparison in the stability in the response between the ECAL energy measurements with no correction, only EFlow correction and EFlow and E/P correction. Using data from the 2018 data taking period.

2.8 Combined Calibration

2.8.1 Structure of Calculation Workflow

Since its measurement of the changing energy scale comes from the flatness of the energy distribution across events and an equalisation of the average energy measurements, this provides less precision in following the regional changes resulting from PN ageing than the E/p method but it provides per crystal granularity on the response where E/p only works at the level of the harness region. These two can be combined as shown by [C$^+$22].

The E/P calculations are dependent on the ecalelf-ntuple- wskim and zskim tasks, so for the recalculation of the calibration constants including the EFlow

calibration the ecalelf ntuples must be recalculated making use of a new calibration of the ECAL informed by the calibration constants found in the EFlow task, following equation 2.3. In the automation framework the resulting workflow would look like figure 2.7.

2.8.2 Example Results from Manual Execution

Previously [C+22] manually executed calculations performing this. The Z boson rest energy measurements are used to monitor the regional energy scale after the combination of: no calibration, EFlow calibration only and E/P calibration after the EFlow. This checks that the energy measurement of each crystal is constant in time with the results of this being shown in figure 2.8.

As would be expected this shows that EFlow decreases the rate of response loss, but does not remove it, and the EFlow + E/P calibration removes it entirely. The EFlow provides the per crystal on it's own shown in figure 2.8, so those channels of the detector which have a particular change in energy scale receive some correction to be closer to the rest of the detector. This indicates that the calibration is working meaning it is sufficient (although additional work may make the calibration more precise), therefore automation of this process is sufficient for real time ECAL calibration. Additional intercalibration methods as outlined in section 1.4 may improve the accuracy of the ECAL calibration.

Chapter 3

Results

3.1 Automation Campaign Example Results

Some test results using the automation framework outlined above are shown below. These are obtained from the data corresponding to the CMS run between 384000 and 386800, performed in a test campaign where only the tasks relevant to the EFlow calculations are performed (see figure 2.5). This is two months of LHC data from 08/03/24 to 10/10/24, making up approximately 1/3 of the 2024 run 3 data. The data grouping used required 2000 pb^{-1} of data from a combination of runs for each Job. When fully implemented, the system will execute these calculations as the data is collected and the required prior tasks are completed.

3.1.1 Weights

The weight distribution over the η axis for three example blocks of data are shown in figure 3.1. The similarity between the distribution for each data group and the symmetry around the $i\eta = 0$ value are expected given $\omega(\eta)$ is primarily a reflection of the detector design and (except for dead channels) the CMS ECAL is symmetrical between positive and negative η. See section 2.7.1 for a further discussion and analysis of how the size of the chosen dataset affects the results of the weights calculations.

3.1.2 κ Values

Figure 3.2 shows the maps of the κ values found by the method outlined in section 2.4. Following the work of [fla23] the κ values were re-calculated for each of the data groups. In figure 3.2 there is a small systematic change in the distributions. The simplest example would be the appearance of a new group of dead channels around $i\phi \approx 50$, $65 < i\eta < 70$ between the runs

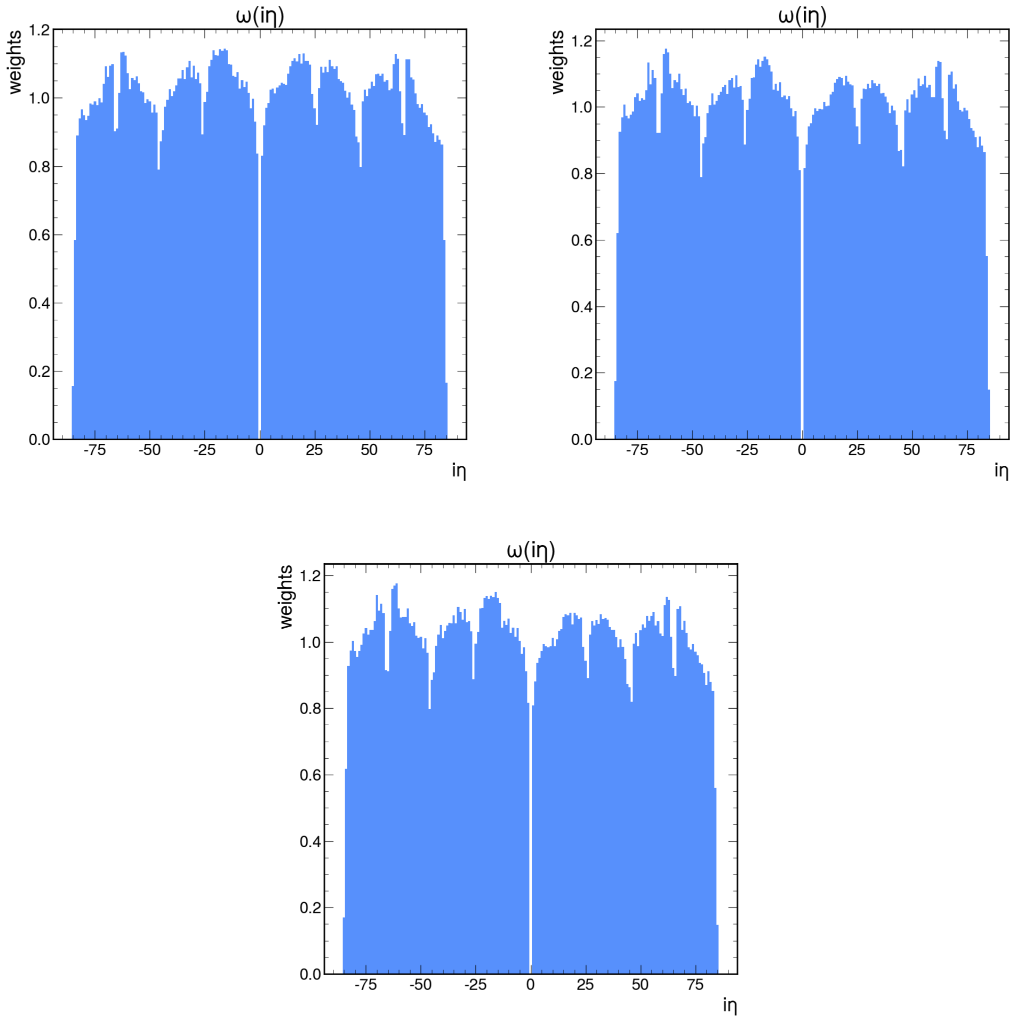

FIGURE 3.1: Here are the weight distributions $\omega(\eta)$ calculated for each data group. The x-axis variable $i\eta$ is the crystal number away from the centre. The results for these 3 of the data groups calculated in this test campaign have no visible systematic difference consistent with the understanding of the $\omega(\eta)$ distribution as a feature of the detector. The plots above are for the data groupings ending in the runs 384069 (top left), 386673 (top right) and 386704 (bottom).

385194 and 386614. It also appears that the κ values are generally decreasing over time. Given how κ is calculated, this is be because as the detector is irradiated a smaller proportion of the energy which is outside the selection window is below it and a greater proportion is above. See figure 2.2 for two examples of these energy distributions. It will be seen later that this leads to a general increase in the values of the intercalibration constants over time.

If the average value of κ over all the crystal channels is found for each data group the bulk changes look like figure 3.3. Looking at the distributions in figure 3.4 there are two peaks in κ frequency at around 0.5 and 2.4, the location of which is fairly constant however the relative height of which changes over time during data taking with the lower one growing and the smaller one shrinking so the average decreases over time. It is not known

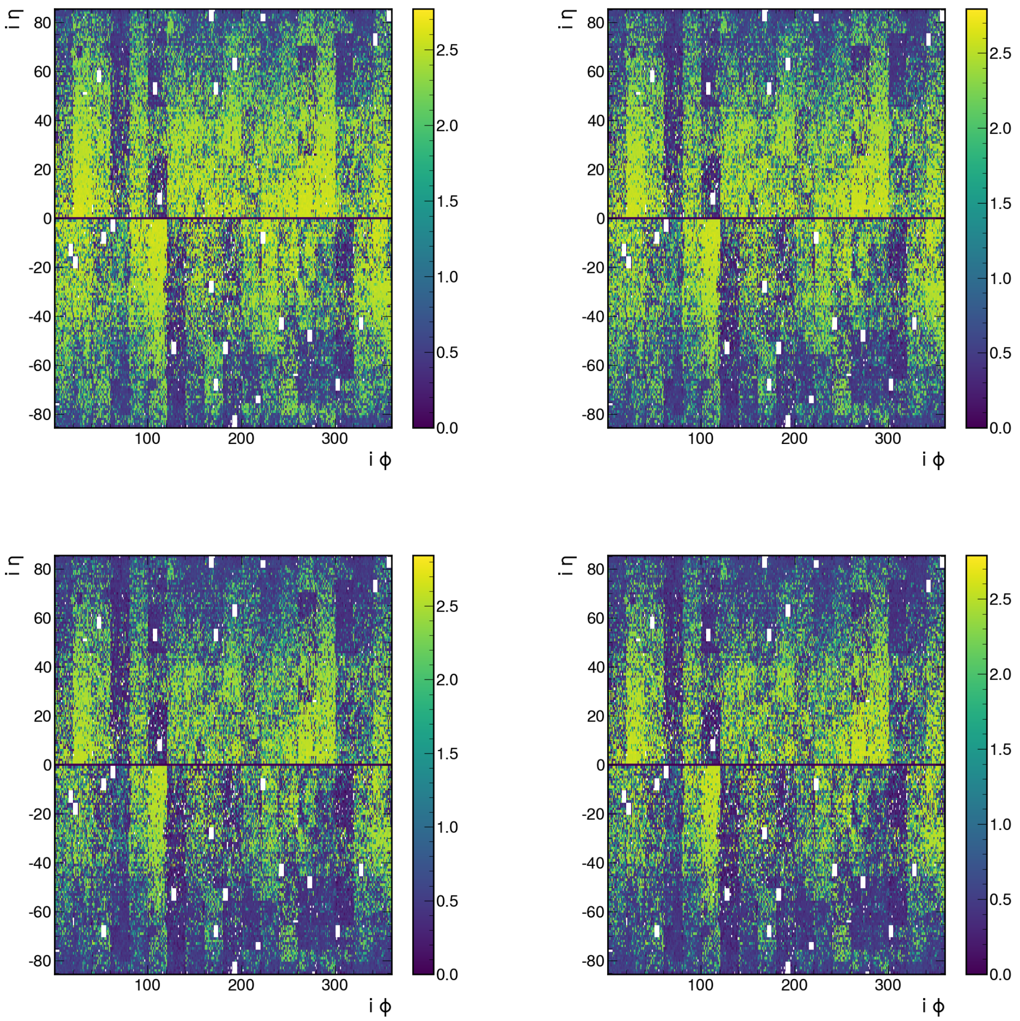

FIGURE 3.2: Maps showing the κ values for each crystal across the barrel of the detector. Results are for data groups ending in the runs 384069 (top left), 385194 (top right), 386614 (bottom left) and 386673 (bottom right).

at this stage why this happens, one way to shed some light on this would be to find the change in κ between years, as this may indicate the relevance of the radiation darkening processes outlined in section 1.3. It would also be valuable to investigate whether the crystal κ values are jumping from the right to left peaks in figure 3.4 or if they're smoothly transitioning across the gap between them.

As can be seen from figure 3.5, over the shutdown period between the 2023 and 2024 data taking there was a recovery of the channels so the frequency peak at $\kappa \approx 0.5$ is significantly smaller after the shutdown. Given the crystal transparency recovers from the electromagnetic radiation effects but not the hadronic, the transfer of crystal channels to the lower κ peak is primarily or entirely a result of electromagnetic radiation.

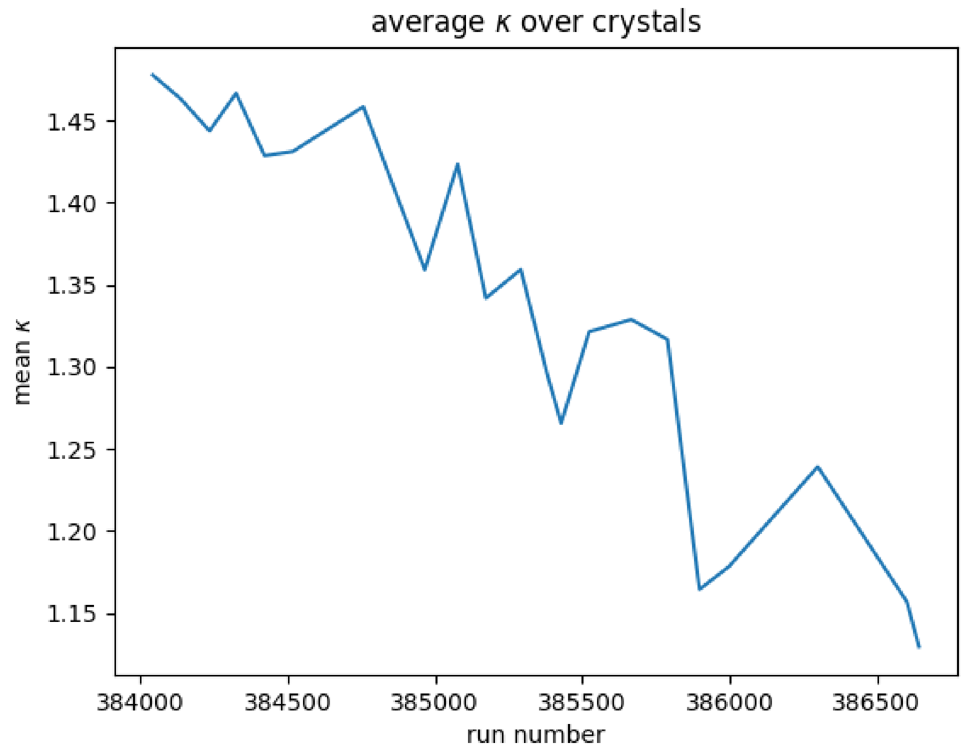

FIGURE 3.3: The κ for each data group averaged over all the crystals in the barrel and plotted against the central run number of the group.

The κ histories for individual crystals in figure 3.6 show that for those channels which are decreasing (all are on average decreasing or constant) do so smoothly through the values between the frequency peaks, then stay constant at around 0.5 rather than jumping down. This is consistent with the continuous effects of the irradiation on the detector.

3.1.3 Calibration Changes over Time

Some examples of how the intercalibration constants change between data periods are shown in figure 3.7. The data periods here are broken up as with the weights above so that each intercalibration has $\approx 2000pb^{-1}$ of data. The appearance of a high level of intrinsic dispersion would suggests that (assuming there aren't rapidly changing effects influencing the intercalibration) consistent results require a larger data group, except the following optimisation study (section 3.2.2) suggests that many of these seemingly random movements are actually systematic changes, the tracking of which precise calibration requires. In particular these smaller changes in individual crystal calibration are the feature of the EFlow calibration in in it's capacity to follow the precise changes in calibration of individual crystals.

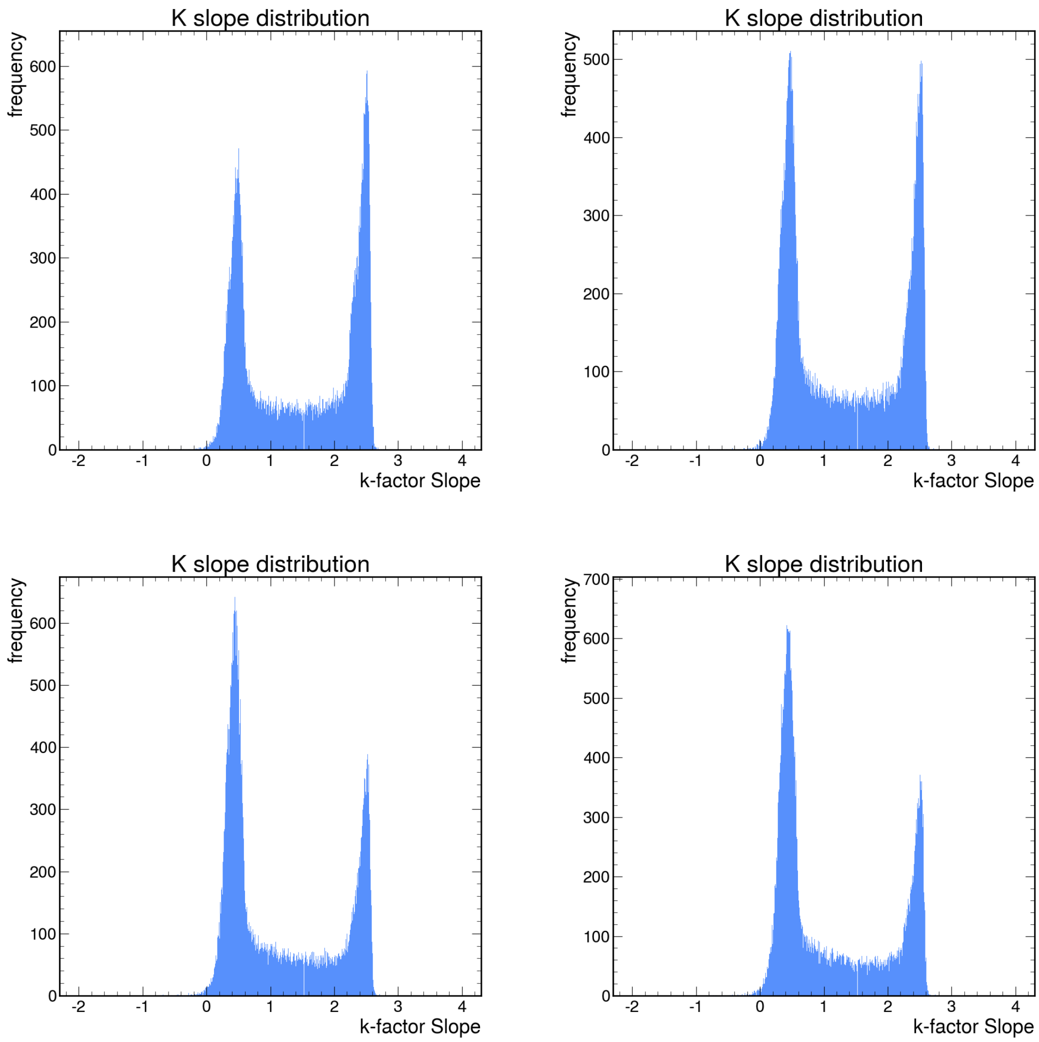

FIGURE 3.4: The distributions of the κ values calculated per crystal. Each distribution is for a different data group with them ending in the runs 384069 (top left), 385194 (top right), 386614 (bottom left) and 386673 (bottom right), the same as for figure 3.2.

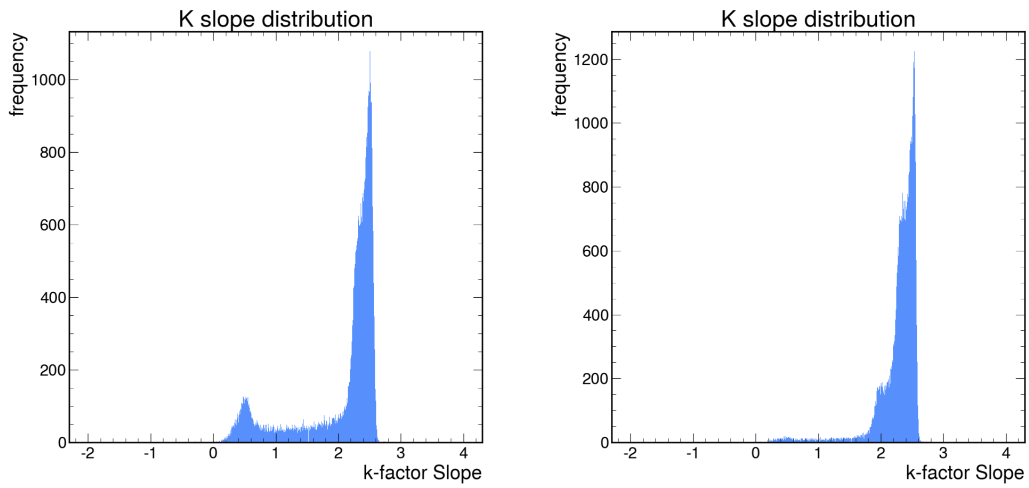

FIGURE 3.5: Two kappa distributions equivalent to in figure 3.4. [Left] is from September 2023 (run 370580) and [right] is from April 2024 (run 279356) These data selections are from the end and start of the 2023 and 2024 LHC periods respectively.

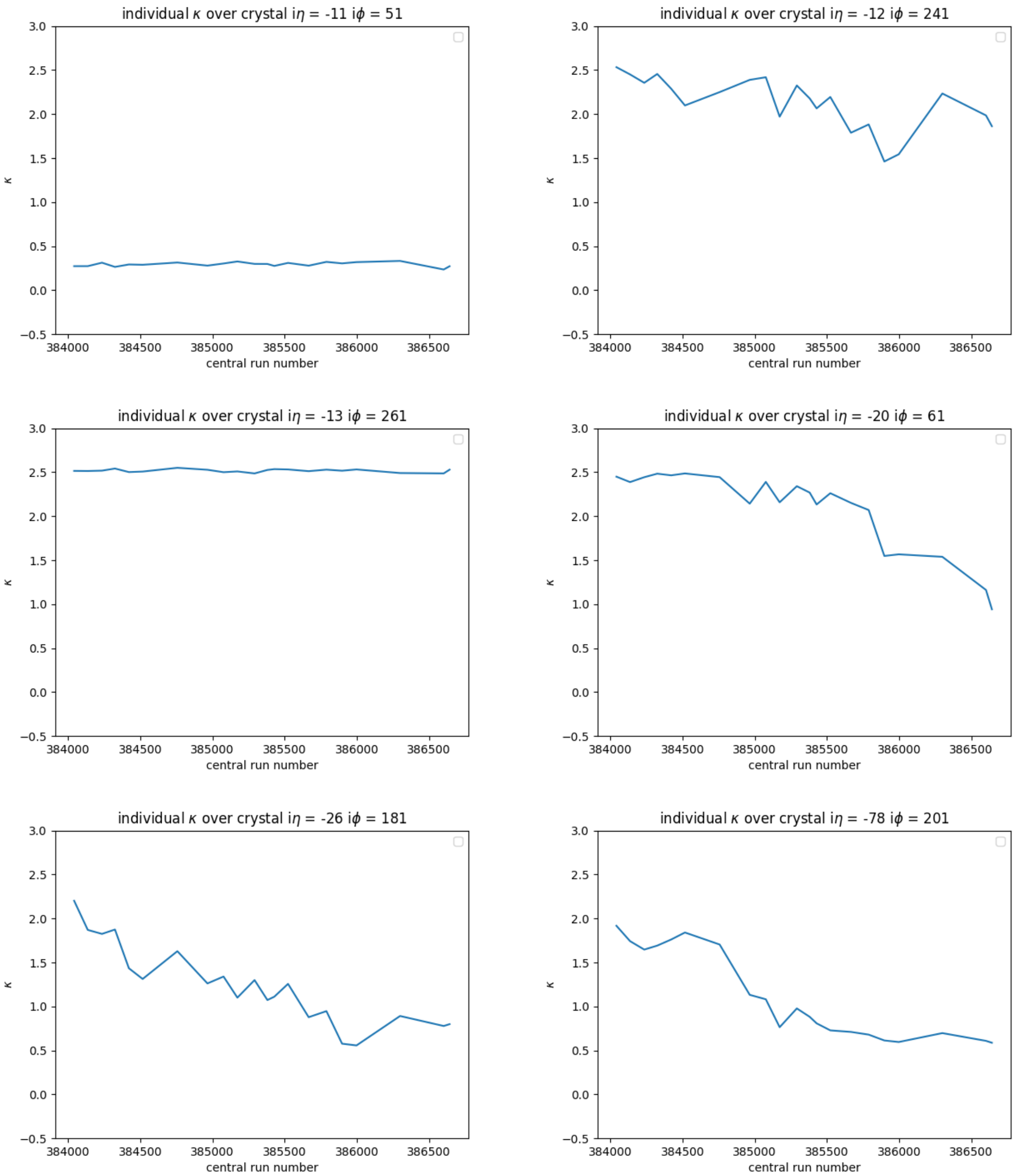

FIGURE 3.6: A sample of the progress of the progress of the κ values of individual crystal channels over the period used for the test campaign.

As is expected given the result in figure 3.3 the overwhelming majority of crystals show a decrease in the inter-calibration. From the gradient map shown in the top right of figure 3.8 it can be seen that there is some supermodule dependence in the gradient as the extended white (indicating no or little change) regions roughly follow the supermodule regions however this pattern is not precisely followed. The peaking of the fit intercepts is an expected implication of the normalisation of the histories to the intercalibration values of the first data group.

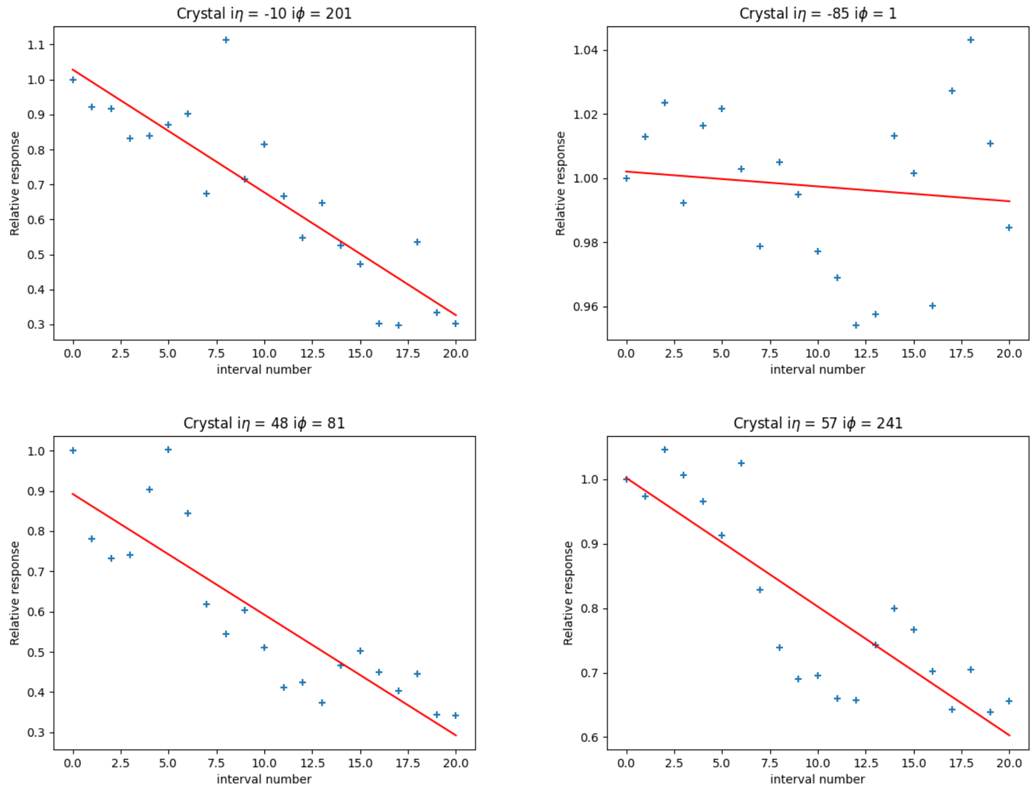

FIGURE 3.7: Some examples of the EFlow intercalibration results for the crystals at the specified η and ϕ locations. The EFlow weighted points are the IC_{ch} values calculated with equation 2.4. A linear fit is found for each crystal history to illustrate something of the general trend.

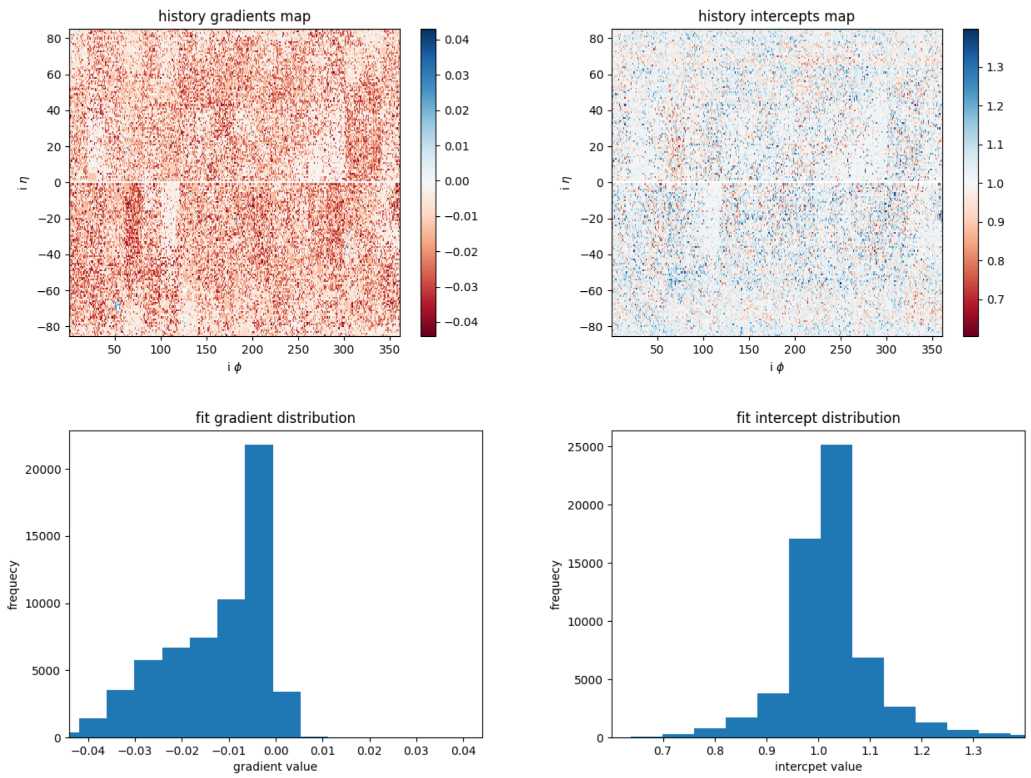

FIGURE 3.8: A set of plots showing the aggregate results of the linear fits in figure 3.7.

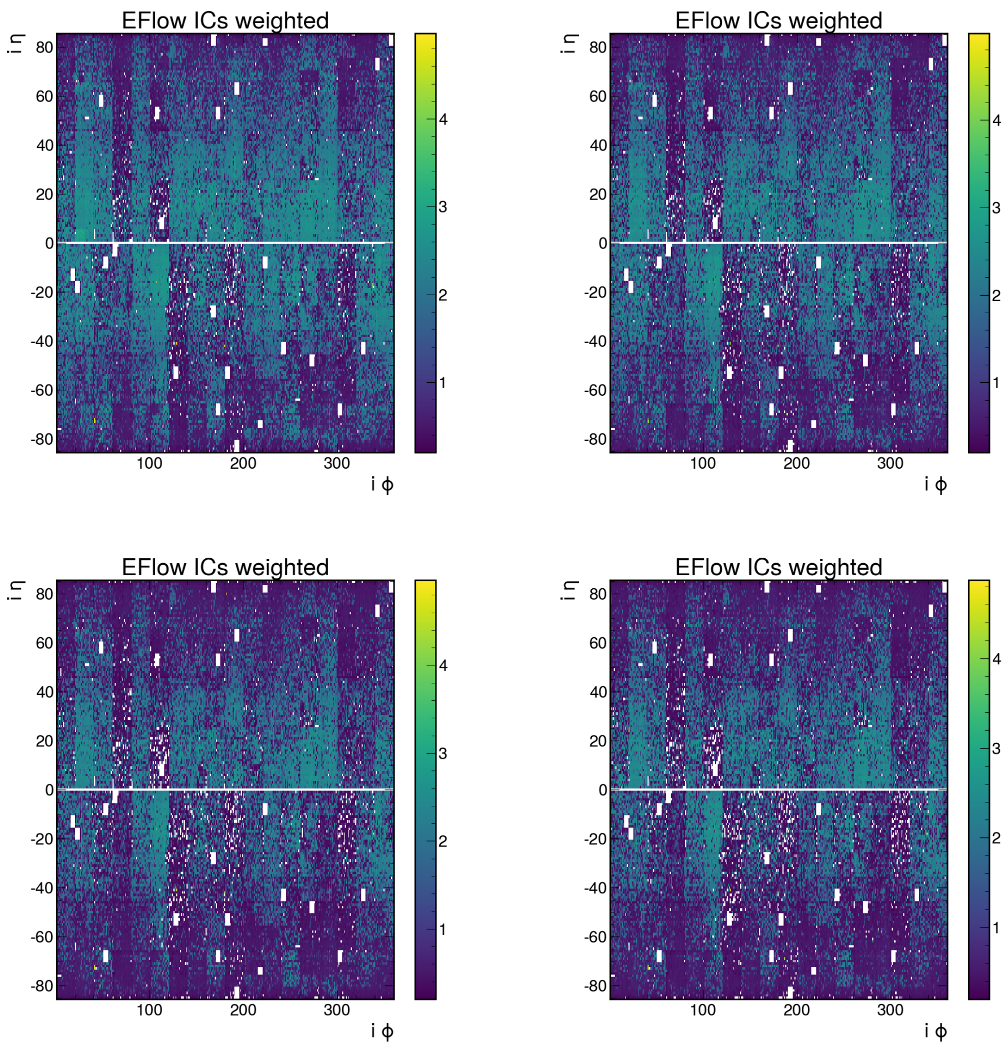

FIGURE 3.9: Two-dimensional maps of the (non-normalised) intercalibration constant values for data groups ending in the runs 384069 (top left), 385194 (top right), 386614 (bottom left) and 386673 (bottom right), matching figure 3.2.

3.1.4 Detector Maps

Figure 3.9 shows four two-dimensional maps of the (non-normalised) intercalibration constant values. Calculated for data groups through the data used for the test campaign. The primary features of these plots are:
(1) The supermodule structure of the ECAL is visible more extensively and clearly in the map of the absolute intercalibration constants than for the normalised constants and with the structure being, for the most part, clear in the same places as it is for the κ maps in figure 3.2. This indicates that the spacial dependence of κ is significantly cancelled out during the normalisation.
(2) A new group of dead channels appears in the bottom left of the map around $i\eta = -69$, $i\phi = 53$.
(3) The intercalibration values decrease over time as expected given the observed change in κ. This is the same as to the results found by [C+22]

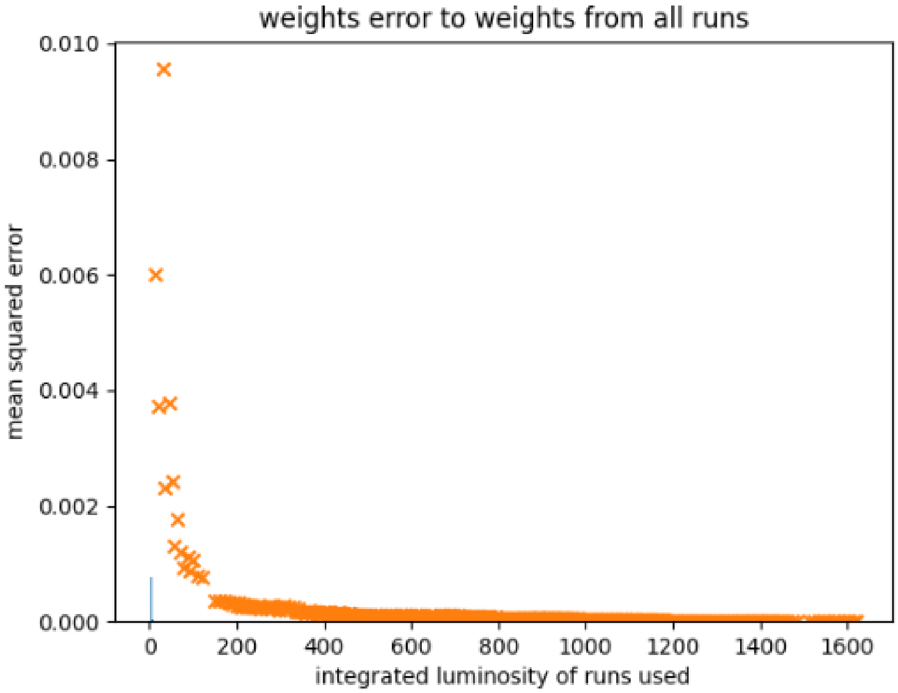

FIGURE 3.10: The relative error between weights values calculated for all the possible subsets of the four runs under consideration in these test calculations and the weights calculated from the data for all 9 runs: 386616,386617,386618,386629,386630,386640,386642,386661,386668.

(pg51), who saw a constant average change in energy response (using 2018 data), noting that EFlow (like ϕ symmetry) is blind to the average changes in the energy response across the barrel region. Therefore it is not guaranteed that κ, which is intended to correct for asymmetries in the effects of miscalibration when applying the selection window, provides a measure of the scale of the ECAL although it may because it does provide a measure of the flatness of the energy distribution of the crystal. The absolute scale of the energy response, in the current workflow, is corrected by comparing Monte-Carlo simulations of the decay of the Z bosons to the measurements of these features. In future workflows this could be either replaced or combined with an update of the E/p calibration to look at the shift in the energy scales as is laid out in section 2.8.

3.2 Optimisation

3.2.1 Weights

The results for the relative error calculations explained in section 2.7.1 is shown in figure 3.10. The full set of run combinations used is all the possible combinations of the runs available, because there is very little change in the weights values over time, so the weights values that in the ideal case should

be calculated from every point within this data set should be almost identical. From figure 3.10, the results are as expected, that the error decreases with increasing integrated luminosity. The decrease in relative error with increasing integrated luminosity over the set of runs in question confirms the expectation that increasing data sample size decreases the error size with diminishing returns, so sample size is a source of error. Diminishing returns in the decrease of the relative error starts at 800 pb^{-1}, therefore, the target luminosity of the data grouping should be at least this high.

3.2.2 EFlow

Four example crystals showing the non-normalised EFlow calibration constants calculated with varying data periods as described in section 2.7.2 can be seen in figure 3.11 with the calibration points plotted against the run number at the centre of the point's data group. Reassuringly, for all of the crystals displayed, except 43200, the intercalibration constants tightly move together; this is true for the overwhelming majority of the crystals. Only the ICs calculated with a group target luminosity of 1000 pb^{-1} consistently show a significant level of intrinsic dispersion away from the mean, and those at 1500 pb^{-1} occasionally do. Therefore a target integrated luminosity of at least 2000 pb^{-1} should be used. For the regions where there is more rapid change in the intercalibration, a data grouping of 3000 $pb^{-}1$ leads to to much of a a delay in the updating of the intercalibration.

The ratios between the mean squared error of the Square Distribution fits and Gaussian Distribution fits to each histogram of the intercalibration constants are the error ratios in the top left of each plot in figure 3.11. Calculated by the equation:

$$Err_{ratio} = \frac{\sum(\rho_{square}(IC) - \rho_{data}(IC))^2}{\sum(\rho_{gauss}(IC) - \rho_{data}(IC))^2} \tag{3.1}$$

Where the summations are along the IC axis, ρ_{data} is a normalised frequency distribution to of the intercalibration constants for a particular data group and ρ_{square} is the best fit of a square distribution and ρ_{gauss} is thee best Gaussian fit. A low error ratio indicates that the square distribution is a better fit than the Gaussian so the changes in the intercalibration are more a systematic change and less a result of random noise. However, this is a reliable tool to the extent that the systematic change in the IC values over time are approximately linear because this is the change that would produce a square distribution. Of the plots shown, the only one that obeys this requirement is the plot of crystal 43200.

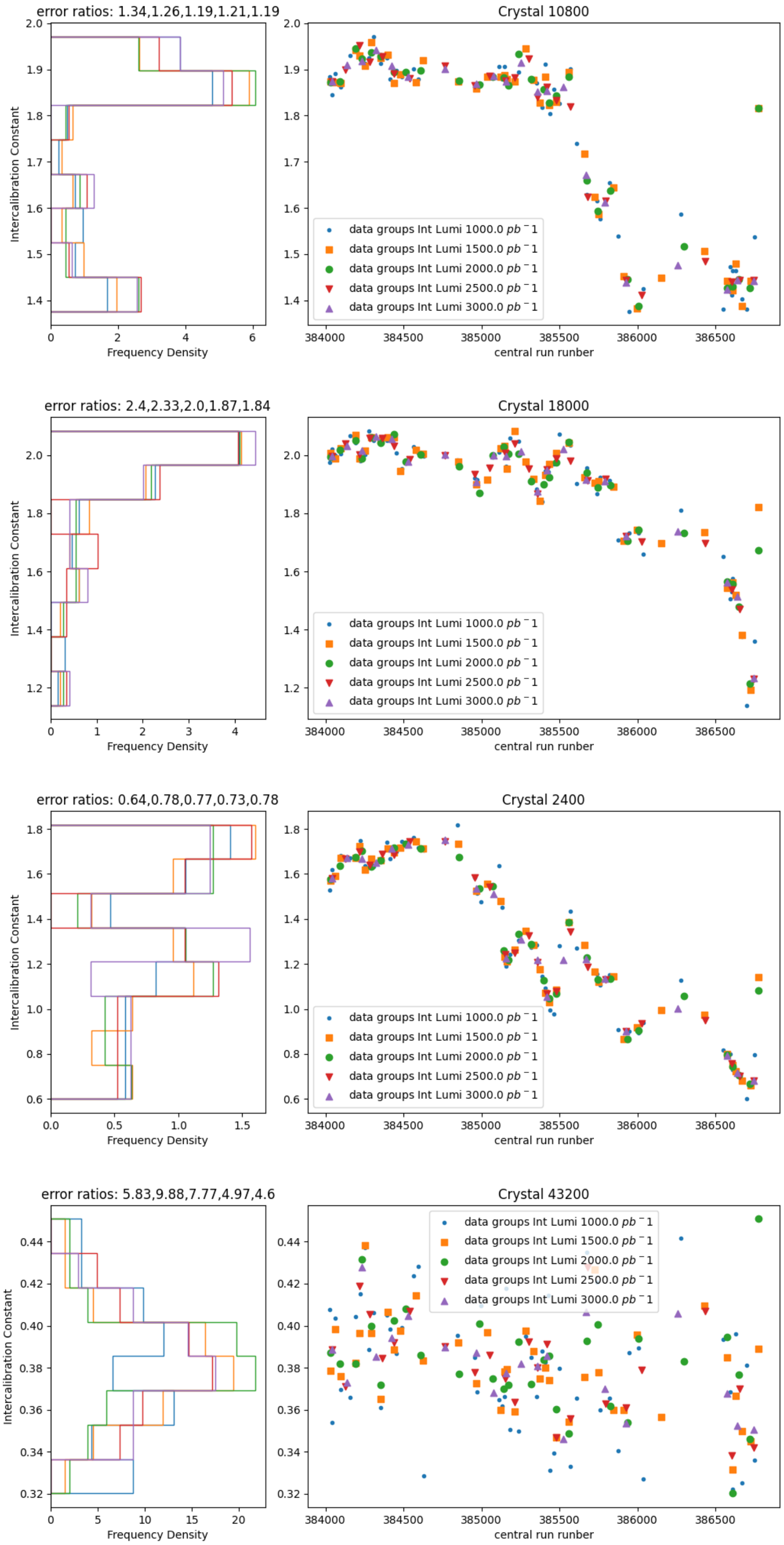

FIGURE 3.11: EFlow intercalibration constants plotted for the portion of the run 3 data from 03/08/2024 until 10/10/2024. They are calculated with varying target data groupings as indicated on the plot. The vertical histogram on the left shows the frequency density of the intercalibration values by target integrated luminosity. The error ratios in the titles of the left plots are the ratios of the mean squared error of a square distribution fit and Gaussian fit to the plotted histogram of the intercalibration constants.

The first thing that is clear from the plots in figure 3.11 is, since there is more variation between the error ratio values for each crystal plot than between the values for different data group size options on the same crystal, that the error ratio is more dependent on the fundamental data source than it is on the specific choice of group luminosity made in data processing. This is unsurprising given the specific requirement of the square distribution plot. If this error ratio is a good indication of the extent to which a set of IC values is influenced by noise, then it would be expected to decrease for increasing data group sizes. For the most part this happens however even on an individual crystal basis the pattern is not consistent except that the decrease is most prominent between 2000 $pb{-}1$ and 2500 pb^{-1} suggesting that using a data grouping target of at least 2500 $pb{-}1$ may yield more accurate results.

A compromise must be reached so that the crystal channels with rapid systematic changes are not updating their calibration too infrequently, and for noise-dominated crystals, there is enough data per intercalibration to minimise the non-systematic changes in the intercalibration. Given that there are significantly more of the first type of crystal channel, a target integrated luminosity of 2000 pb^{-1} is chosen; this is also computationally convenient because it is the same data group length as the E/p calibration. A lower group size would also likely not be useful even if it did provide a more precise calibration, since it is only feasible to update the calibration every three or so days, which, during peak data taking, gives about 2-3 fb^{-1} depending on the performance of the LHC. If a greater proportion of the crystal channels had the noise level of crystal 43200, then a higher target integrated luminosity would be preferable.

Chapter 4

Conclusions and Further Prospects

The results of the Example workflow shown in section 3.1 demonstrate the applicability of the developed workflow to large amounts of CMS data which will translate to running as part of the production campaign during data collection in 2025 and beyond. The features of the results match those presented by [C$^+$22]; however, the computational process is more general, and the move from data being grouped by integrated luminosity rather than number of hits improves the consistency of the data grouping. The completion of the calibration is done by comparing Monte-Carlo simulations of the $Z \to e^+e^-$ decays to the measured results to provide an absolute scale.

The results in section 2.7.2 verify the choice to use an integrated luminosity grouping of $2000pb^{-1}$. This can be seen in that for the crystals with a greater noise influence on the intercalibration constant calculation, the movement of the intercalibration constants do not track together when the data groups contain less than 2000$pb{-}1$. For the noise heavy crystals it would be ideal to have a data grouping of at least $2500pb^{-1}$, however for the crystals where the systematic change is faster this would lead to the change between intercalibration constants being too great, so much of the data would be processed with IC values which are no longer applicable because the detector has changed. $2000pb^{-1}$, which is the same data grouping as was chosen for the E/P calibration development, is therefore chosen as a compromise.

The combining of the E/P and EFlow calibrations outlined in section 2.8 was not completed due to time constraints. This is the next step in improving the accuracy of the automated calibration of the CMS ECAL. The E/P intercalibration is not able to provide accurate per crystal corrections to the ECAL, which the EFlow is, so the two combined are necessary for complete automated correction for the PN ageing error on the laser calibrations. Once

this is complete, adding additional calibration techniques to the workflow will improve the precision of the calibration; however, these must not use the same CMS data and physics as the techniques that are used to monitor the accuracy of the calibration, or the calibration and monitoring errors will not be independent.

There is also room for further study of the changing κ distribution, resulting from the changing energy distributions. It is unclear what the physical origin of the decreasing average κ is. The found results are consistent with electromagnetic radiation effects flattening the energy distribution on the channels, decreasing the average κ, however it would be helpful to have a clear mechanism for how the radiation effects flatten the energy distribution. This may provide insight into a more subtle characteristic in how the radiation damage incurred during data taking influences the detector, and may validate the efficacy of EFlow in accounting for the residual PN ageing.

Bibliography

[4le14] Measurement of the properties of a higgs boson in the four-lepton final state. *Physical Review D*, 89(9), May 2014.

[AAK+98] Alexander Annenkov, E Auffray, Mikhail Korzhik, P Lecoq, and J-P Peigneux. On the origin of the transmission damage in lead tungstate crystals under irradiation. *physica status solidi (a)*, 170(1):47–62, 1998.

[AB21] Douglas Ross Alexander Belyaev. *The Basics of Nuclear and Particle Physics*. Springer, Undergraduate Texts in Physics, 2021.

[AC16] T. Adams and CMS. Beam test evaluation of electromagnetic calorimeter modules made from proton-damaged pbwo4 crystals. *Journal of Instrumentation*, 11(04):P04012, apr 2016.

[aut25] Automation gitlab repository. `https://gitlab.cern.ch/cms-ecal-dpg/ECALELFS/automation`, 2025.

[B+20] A Beschi et al. *Measurement of the associated production of top quark pairs with a Higgs boson in the diphoton final state with the CMS detector*. PhD thesis, 2020.

[C+21] CMS et al. Electron and photon reconstruction and identification with the cms experiment at the cern lhc. *Journal of Instrumentation*, 2021.

[C+22] F Cetorelli et al. *First observation of the electroweak production of a W+ W-pair in association with two jets in the fully leptonic channel with the CMS detector at the LHC*. PhD thesis, 2022.

[Cc+24] Giacomo Cucciati, CMS collaboration, et al. The upgrade of the cms electromagnetic calorimeter for hl-lhc. *Journal of Instrumentation*, 19(05):C05035, 2024.

[CMS97] *CMS The Electromagnetic Calorimeter Project, Technical Design Report*. 1997.

[CMS17] The Phase-2 Upgrade of the CMS Barrel Calorimeters. Technical report, CERN, Geneva, 2017. This is the final version, approved by the LHCC.

[CMS21] Cms ecal calibration software. `https://github.com/cms-sw/cmssw/tree/master/Calibration/EcalCalibAlgos`, 2021.

[cms24] Performance of the cms electromagnetic calorimeter in pp collisions at s = 169 tev. Technical Report 9, CERN, 2024.

[CMS25] CMS. Review of top quark mass measurements in cms. Technical report, 2025.

[DGV+04] F Drouhin, L Gross, D Vintache, A Marchioro, C Paillard, C Ljuslin, P Siegrist, A L Tsirou, L Mirabito, and P G Verdini. The CERN CMS Silicon Strip Tracker Control System. Technical report, CERN, Geneva, 2004.

[eca25a] Ecalphisym gitlab repository. https://gitlab.cern.ch/cms-ecal-dpg/ECALELFS/ecalphisym, 2025.

[eca25b] Cms ecalautoctrl docs. https://cmsecaldocs.web.cern.ch/cmsecaldocs/ecalautoctrl/, Accessed: 17/01/2025.

[fla23] Flavia cotrelli's ecalphisym github repository. https://github.com/flaviacetorelli/ecalphisym, 2023.

[GAB+99] K. Gill, C. Azevedo, J. Batten, G. Cervelli, R. Grabit, F. Jensen, J. Troska, and F. Vasey. Ageing tests of radiation damaged lasers and photodiodes for the cms experiment at cern. In *1999 Fifth European Conference on Radiation and Its Effects on Components and Systems. RADECS 99 (Cat. No.99TH8471)*, pages 429–436, 1999.

[GC19] Saranya Ghosh and Anna Colaleo. Highlights from the compact muon solenoid (cms) experiment. *Universe*, 5:28, 01 2019.

[gra25] Grafana. https://grafana.com/, Accessed: 17/01/2025.

[HCA97] *The CMS hadron calorimeter project: Technical Design Report.* "Technical design report. CMS". "CERN", 1997.

[inf25] Influxdb. https://www.influxdata.com/, Accessed: 17/01/2025.

[jen25] Jenkins. https://www.jenkins.io/, Accessed: 17/01/2025.

[las19] Cms ecal response to laser light. Technical report, CERN, European centre for Nuclear Research, 2019.

[Pig23] Simone Pigazzini. Automatic data processing for prompt calibration of the cms ecal. Technical report, 2023.

[Rog10] Christopher Rogan. Laser monitoring for the cms ecal. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 623(1):267–269, 2010. 1st International Conference on Technology and Instrumentation in Particle Physics.

[SC22] Uttiya Sarkar and CMS. Searches for supersymmetry in cms. *International Journal of Modern Physics A*, 2022.

[WLG25] Worrldwide lhc computing grid. https://wlcg.web.cern.ch/, 2025.

[Won94] Cheuk-Yin Wong. *Introduction to High-Energy Heavy-Ion Collisions.* 1994.

Software developed for this project can be found in the following CERN internal GitLab repositories [eca25a] and [aut25]. The software developed prior to this project on the EFlow calibration can be found in the git repository [fla23] and some background environment software is in [CMS21]. All of the plots produced for the test campaign can be seen at https://ecal-automation-plots-dev.web.cern.ch/test_eflow_v1/eflow/.