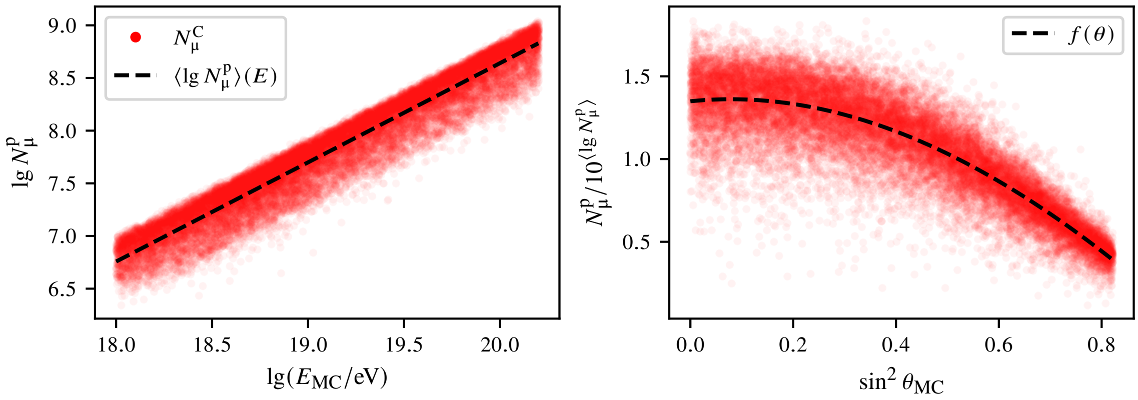

Figure 1: Representation of two-step fit to parametrize $\langle N_\mu^p \rangle$ (see Eq. (2)) from N_μ^C. *Left:* Dependence of N_μ^C on the Monte-Carlo energy E_{MC} for $\theta \in [0°, 65°]$. The average number of muons follows a power law. The black, dashed line represents the best-fit of the exponential in Eq. (2). *Right:* Dependence of relative number of muons on the Monte-Carlo inclination angle θ_{MC}. The more matter is traversed by the shower, the more muons are lost due to attenuation effects. The black, dashed line shows the best-fit of the $f(\theta)$ function.

Table 1: Best-fit parameters of the two-step fitting process of Eq. (2). In both steps (see Fig. 1) least-square fits have been used. The first row contains the best-fit values and the second row the uncertainties of these values.

	A	B	a	b	c	d
$\langle \cdot \rangle$	−10.190	0.942	1.349	0.317	−2.050	0.299
$\sigma_\cdot$	0.033	0.002	0.004	0.046	0.131	0.105

Additional information

This document gives additional information on the fitting process (see Fig. 1 and Table 1) of

$$\langle N_\mu^p \rangle(E, \theta) = f(\theta)\, 10^{A + B \lg E} = [a + bx + cx^2 + dx^3]\, 10^{A + B \lg E} \tag{2}$$

and the NN architecture (see Fig. 2). The architecture of the subnetworks N1 and N2 is described in detail in [1]. There, however, N2 is split into N2 and N3 where former is only a subnetwork dedicated to correlate spatial information.

References

[1] S. T. Hahn. PhD thesis (2022), Karlsruhe Institute of Technology (KIT).

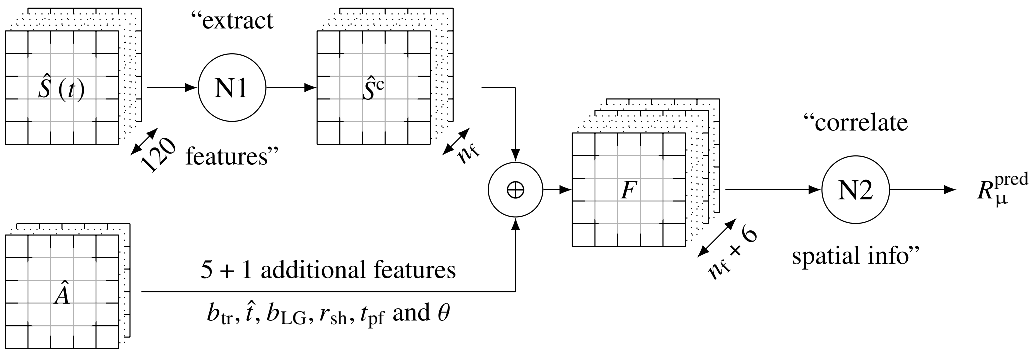

Figure 2: Overview of architecture of the NNs used to infer R_{μ} from shower footprints. The networks have two different inputs, denoted as $\hat{S}$ and $\hat{A}$. The normalized time signals of the triggered detectors $\hat{S}$ are "compressed" in subnetwork N1. Due to weight-sharing, all encoded time signals are treated the same. From 120 time bins to n_{f} (feature-)channels $\hat{S}^{\mathrm{c}}$. The output is concatenated to the additional station- and event-level input data, denoted as $\hat{A}$. The (feature-)tensor F is then used as the input of the second subnetwork N2. N2 is 2d-convolution-based network that exploits the spatial information reducing for the direct prediction of R_{μ}^{pred}.