Black Hole Production in High Energy Collisions

Saeede Nafooshe

University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica

Abstract

Using the BlackMax event generator, we investigate the possibility of black hole production in proton-proton collisions in LHC energy regimes and at higher energies. One priority of extrapolating to the higher energies is to investigate whether black holes are produced by high energy cosmic rays or not. Although black hole signals have not yet been seen, possibility of the existence of large extra dimensions, means that there is hope to see some black hole production signals either in the next run of the LHC or in high energy cosmic rays.

Keywords:
Black hole production, BlackMax event generator

1. Introduction

In the context of black hole production, typical required energies are energies which are greater than the Planck mass i.e. 10^{19}eV. It is impossible to access such a high energy in our world. As this scale is very far from the other fundamental scales like electroweak scale a hierarchy problem arises. However, in 1998 Arkani, Dimopoulos and Dvail (ADD model)[1] suggested a scenario in which this problem has solved. In this *flat extra dimensions*[1] scenario the $(3+1)$-dimensional world that we see is embedded in a higher dimensional universe, so gravity can become strong at a scale around 1 TeV. At this scale, we can expect to produce black holes by high energy colliders or by high energy cosmic rays (which corespondes to collisions at center of mass energy up to 140TeV).

Email address: `saeede.nafooshe@ung.si` (Saeede Nafooshe)

[1]The extra dimensions are not necessarily flat, but could be warped [2] and gravity also can be strong within the warped extra dimensional volume

In general the production of a black hole takes places in four stages: balding phase, spindown phase, Schwarzschild evaporation and Planck phase. Study of each of these stages is very important, as one should apply all these theoretical studies to simulate the black hole production. Having the simulated results, from event generators like CHARYBDIS2 [3] and Black-Max [4] which are currently being used at the Large Hadron Colider (LHC) to look for signatures of black hole production and evaporationone, one can expect the same tracks for particles in the detectors. The main goal of this paper is to use BlackMax, for investigating the black hole production in different scenarios and energies.

2. Black Hole Production

When two high energy particles that behave like shock waves come close together, they will form a trapped surface even before they collide. This region has a high density such that no radiation can escape from it.The possibility of black hole production can be evaluated by the hoop conjecture which states that an apparent horizon[2] forms when and only when, the mass of system is compacted in a region with circumference c

$$c = 2\pi(2r_h),\tag{1}$$

where r_h is the Schwarzschild horizon radius. This statement means that the impact parameter of the collision should be smaller than $2r_h$. The apparent horizon has the geometry of the flat disk, so the geometrical cross section for the black hole formation is

$$\sigma_{b.h.} = \pi((2\mu)r_h)^2,\tag{2}$$

where μ is the center of mass energy of each incoming particle and $(2\mu)r_h$ is the gravitational radius of the D-dimensional Schwarzschild black hole with mass 2μ.

After event horizon formation the balding phase takes place. Through this phase the black hole will shed all its hair (multipole moments) by classical emission of gravitational waves. During this phase the complex shape of

---

[2]The formation of apparent horizon is the sufficient condition for the formation of event horizon [5].

the black hole will relax to a D-dimensional rotating black hole, described by the following Kerr metric [6]

$$ds^2 = (1 - \frac{\mu r^{4-d}}{\Sigma(r,\theta)})dt^2 - \sin^2\theta(r^2 + a^2(\sin^2\frac{\mu r^{4-d}}{\Sigma(r,\theta)}))d\phi^2$$
$$+ 2a\sin^2\theta\frac{\mu r^{4-d}}{\Sigma(r,\theta)}d\phi dt - \frac{\Sigma(r,\theta)}{\Delta}dr^2 - \Sigma(r,\theta)d\theta^2$$
$$- r^2\cos^2\theta d^{d-3}\Omega, \tag{3}$$

where d is the number of spacelike dimensions and,

$$\Sigma = r^2 + a^2\cos^2\theta, \tag{4}$$

$$\Delta = r^2 + a^2 - \mu r^{4-d}, \tag{5}$$

next the spindown phase follows, during which the spining black hole sheds its angular momentum by Hawking radiation. According to [7] the time scale for this radiation emission is

$$\frac{\Delta t}{r_h} = (\frac{M}{M_D})^{\frac{D-2}{D-3}}, \tag{6}$$

where M_D is the Planck mass. In the semiclassical regime, where the black hole has an exact thermal spectrum, the expectation value for the number of particles emitted by the Hawking radiation and superradiation in the mode ω, is given by

$$< n(\omega) > = \frac{\gamma(\omega)}{e^{\frac{\omega}{T_H}} \pm 1}, \tag{7}$$

where $\gamma(\omega)$ is the graybody factor [3], T_H is the Hawking temperature and, the plus (minus) sign describes fermions (bosons). Several studies of the graybody factors of the D-dimensional black hole are available in [8], [9], [10]. The graybody factor is important because it gives the probability of Standard Model particles emission and graviton radiation on the brane or into the bulk. The black hole continues evaporating until it reaches the Planck scale. It is not possible to use the semiclassical approximation for describing the fourth phase of the black hole production, because higher dimensional quantum gravity effects are important at the Planck scale.

---

[3]The probability for an outgoing wave, in the ω mode, to reach infinity. This coincides with the absorption probability, i.e., the probability for an incoming wave, in the ω mode, to be absorbed by the black hole.

3. Black Hole Simulation

BlackMax is one of the black hole event generators which is designed for investigating black hole production at LHC. However, it is flexible enough that one can change input parameters and extrapolate the results to much higher energies than those that the LHC can reach. The codes are written for proton-proton, proton-antiproton and, electron-positron collisions. In this paper we will present results concerning the simulation for proton-proton collisions. In the first step it uses parton distribution function (PDF) sets to calculate the cross section at the parton level. We use the leading order CTEQ6 PDF set for our purpose, but for different sets of PDFs we can get different results [11]. The next step is interfacing the results to a parton shower and hadronization generator PYTHIA [12]. As the output, the generator gives the Standard Model particles with their energy, linear and angular momentum distributions. For the current work, we are interested in the parton level study as some of the elementary particles produced at the parton level will decay before hadronization and they cannot exist after that.

Thanks to the existence of extra dimensions, the Planck scale can be in the TeV range. As figure 1 shows, the total cross section of a non-rotating black hole on a tensionless brane increases with the number of extra dimensions as a function of energy.
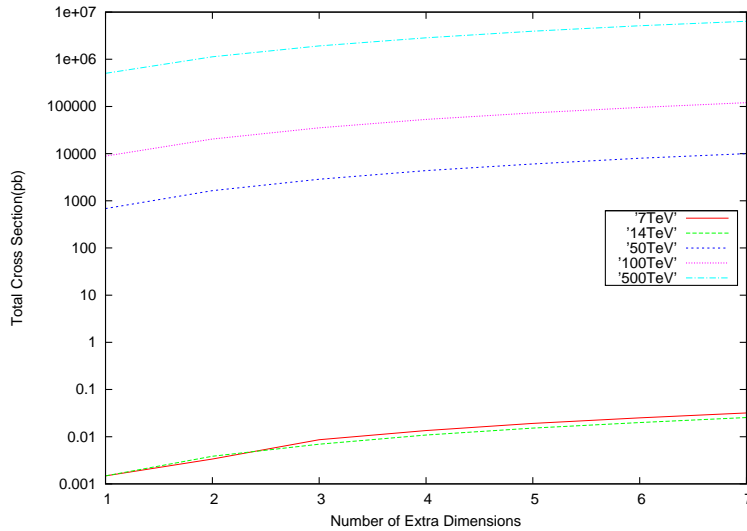

Figure 1: Total cross section for production of black hole for different number of extra dimensions and for different energies.

4

Figure 2 shows the multiplicity of the produced particles with $E_{cm} = 50$ TeV for gauge bosons as well as higgs boson and graviton. After evaporation of the black hole, we have the high number gluons but a small number of gravitons. One of the reasons for this feature is that Hawking radiation is sensitive to the degrees of freedom of particles. It means that for gluons which carries color charges, namely more degrees of freedom, one expects higher multiplicity. However for gravitons, scenario is different. As gravitons can also emit into bulk which this emission decreases their multiplicity in the brane.
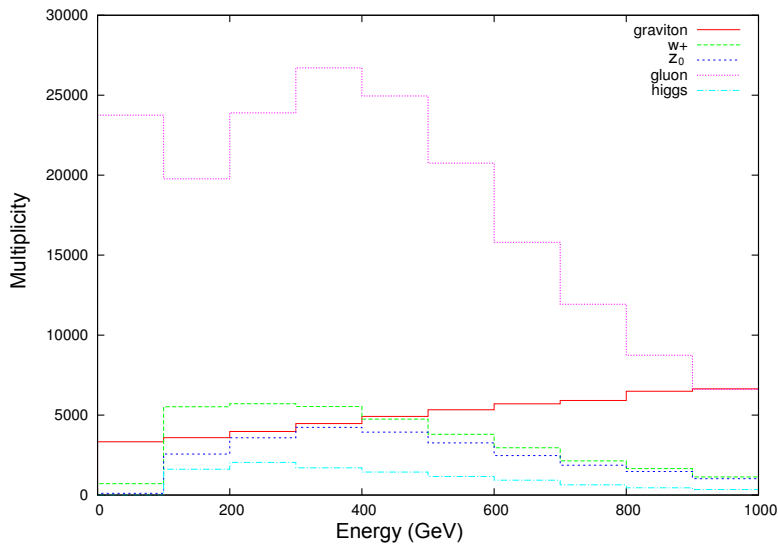

Figure 2: Multiplicity of the produced particles as a function of energy.

4. Conclusions

In this paper, we used BlackMax event generator for black hole production as predicted by the latest theoretical calculations. Flexibility of Black-Max allows us to simulate the black hole production for different scenarios and different ranges of energies. As the energy of the colliding particles increases the probability of producing black holes also increases. The total cross section of black hole production grows with the number of extra dimensions, and as a result the probability of producing a black hole will increase. Here we are concentrating on the Schwarzschild phase which $\sim 70\%$ of the total center of mass energy is radiated. However, for detecting any signals of black hole production, the hadronisation processes

should now be carried out on the initial decay products that we have derived here. For simulation of this step we need to includes PYTHIA codes.

References

[1] N. Arkani-Hamed, S. Dimopoulos and G.R. Dvali, *The hierarchy problem and new dimensions at a millimeter*, Phys. Lett. B429 (3-4) (Jun 1998)263-272, SLAC-PUB-7769, SU-ITP-98/13

[2] L. Randall and R. Sundrum, *A large mass hierarchy from a small extra dimension,* Phys. Rev. Lett. 83 (17) (25 Oct 1999) 3370-3373, MIT-CTP-2860, PUPT-1860, BUHEP-99-9.

[3] C.M. Harris, P. Richardson and B.R. Webber, *CHARYBDIS: A Black Hole Event Generator,* JHEP 0308:033, hep-ph/0307305, 2003

[4] De-Chang Dai, Cigdem Issever, Eram Rizvi, Glenn Starkman, Dejan Stojkovic, Je Tseng *The BlackMax Manual A black-hole event generator with rotation, recoil, split branes, and brane tension,* (2009) arXiv:0902.3577v1 [hep-ph]

[5] R. Wald, *General Relativity* University of Chicago Press, Chicago, 1984, p.310.

[6] R. C. Myers and M. J. Perry, *Black Holes In Higher Dimensional Space-Times,* Annals Phys. 172, 304 (1986).

[7] S. B. Giddings and S. D. Thomas, *High energy colliders as black hole factories: The end of short distance physics,* Phys. Rev. D 65, 056010 (2002) [arXiv:hep-ph/0106219].

[8] P. Kanti and J. March-Russell, *Calculable corrections to brane black hole decay. I: The scalar case,* Phys. Rev. D 66, 024023 (2002) [arXiv:hep-ph/0203223].

[9] P. Kanti and J. March-Russell, *Calculable corrections to brane black hole decay. II: Greybody factors for spin 1/2 and 1,* Phys. Rev. D 67, 104019 (2003) [arXiv:hep-ph/0212199]

[10] D. Ida, K. y. Oda and S. C. Park, *Rotating black holes at future colliders: Greybody factors for brane fields,* Phys. Rev. D 67, 064025 (2003) Erratum-ibid. D 69, 049901 (2004) arXiv:hep-th/0212108

[11] ATLAS Collaboration, *Search for strong gravity signatures in same-sign dimuon final states using the ATLAS detector at the LHC,* Phys. Lett. B. 709(2012) 322-340

[12] T. Sjostrand, S. Mrenna, P. Z. Skands, *PYTHIA 6.4 Physics and Manual,* JHEP 05 (2006) 026

Electron-beam optimization studies for the FERMI@Elettra free-electron laser

Eugenio Ferrari

University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica, Slovenia

Sincrotrone Trieste SCpA, Strada Statale 14 - km 163,5, Basovizza (TS), Italy

Abstract

FERMI@Elettra is a single-pass free-electron laser, based on seeded high-gain harmonic generation. Presently, the first phase of the project (covering the spectral range between 100 and 20 nm) is under commissioning. The free-electron laser performance depends on the quality of the electron beam. In the case of the FERMI linear accelerator, the latter is strongly influenced both by the wake-fields present in the accelerating sections and by possible misalignments of the various accelerator components. In order to investigate and compensate these effects, we performed a study based on local trajectory bumps. We demonstrate that this approach significantly improves the electron-beam quality and, eventually, the free-electron laser performance.

Keywords: LINAC, Wakefields, Free-electron Laser

1. Introduction

An electron beam, usually provided by a linear accelerator (LINAC) or by a storage ring (SR), should have high intensity, small dimensions and reduced spread in energy [1–8] in order to be suitable for most of nowadays applications, like high-brilliance colliders and free-electron lasers (FELs'). One of the most commonly used figure of merit for the beam quality is the (normalized) emittance [9], a quantity that is proportional to the area of the phase space occupied by the beam. It is customary to express it as $\epsilon_i = \Delta i \Delta i'$, where $i = \{x, y\}$, where Δi is the transverse dimension of the beam and $\Delta i'$ is the beam divergence.

Email address: `eugenio.ferrari@elettra.trieste.it` (Eugenio Ferrari)

In the case of storage rings [10] the beam must be kept stable and one of the main concerns is the cope with the instabilities: these can build up coherently if not correctly controlled, resulting in beam degradation and eventually in beam loss. One of the main challenges for linacs is to mantein the horizontal (vertical) emittance preserved along the whole accelerator. The emittance can be degraded along the machine due to instabilities. These can be either shot-to-shot related (jitter) or static systematic distortions [11]. Systematic phenomena that can cause emittance increase are related to the interaction of the electron bunch with its surrounding environment (beam pipes, accelerating structures, discontinuities, etc.), which is referred as wakefields [12, 13], or to the misalignment between accelerating sections and quadrupoles. These effects have been widely investigated for linear colliders in terms of luminosity performance and emittance growth [14–16]. Jitter instabilities are instead related to stochastic phenomena, like vibrations of the magnets, current ripple of power supplies, noise in the phases of the accelerating fields and so on.

In order to prevent the blow-up of the emittance, local trajectory bumps can be applied [17–19]. It is possible to distort the trajectory in order to create a controlled, local misalignment in the machine by using steering magnets. These bumps are used to reduce the emittance dilution because they introduce a well defined and controllable dispersion[1] in the beam, which can be set in amplitude and betatron phase, in order to compensate the instabilities and misalignments introduced in previous parts of the accelerator. This technique is empirical and determines a "golden trajectory" for the machine. In this paper we present an implementation of the local-bumps method by means of trajectory variations around the accelerating cavities. It was performed in order to both reduce the misalignment errors of the accelerating structures and to investigate the effects of the geometrical wakefields[2] on the electron beam. The study here presented was carried on at the FERMI@Elettra linac [21]. Fermi is a 4th generation lightsource, based on a single-pass seeded FEL scheme. Free-electron lasers' are powerful sources of coherent radiation in a wide range of the electromagnetic spectrum. The working principle of an FEL relies on the interaction between an ultra-relativistic electron beam and a co-propagating electromag-

_____

[1]Dispersion is the momentum dependence of charged particle deflections in a magnetic field [20].

[2]This kind of wakefields are generated by the interaction of the head of the electron bunch with its surrounding, which is reflected back to the beam itself, and interact with the tail of the same bunch

netic wave, in the presence of a static and periodic magnetic field, usually generated by an undulator [22, 23]. This interaction happens usually in a single passage of the electrons through the insertion device, so this kind of lightsources are usually referred to single-pass FELs. Depending on the origin of the electromagnetic wave which co-propagate with the electron beam inside the undulator, one can distinguish between SASE (Self Amplified Spontaneous Emission), if the electromagnetic field is generated by the spontaneous emission of the electrons, and Seeded configuration where the wave is provided by an external source (usually a conventional laser). After presenting the FERMI@Elettra linac configuration, we will focus on the description of the tools needed to implement the method. We will show the results we experimentally obtained and compare them with simulations of the machine.

2. Method

In order to evaluate the correct alignment of the accelerating cavities with respect to the BPMs' we performed trajectory scans. Using the trajectory feedback capabilities [30], we are able to vary the offset of the trajectory at different positions of the machine. Trajectory bumps can be performed around one or more accelerating structures composing the linac (cfr. fig. 1). To do so we used one or more BPMs' around each linac section.

2.1. Linac description

The FERMI@Elettra linac provides the electron beam necessary for the operations of the FEL radiation source. It accelerates the 4 MeV electrons produced by the photo-cathode (injector) up to 1.2 GeV (see fig. 1). It is composed of two different types of accelerating structures:

- seven, 4.5-m-long SLAC-type, constant gradient structures (linac 1 and linac 2), with maximum energy gain per cavity 47 MeV;

- seven, 6.1-m-long Backward-Traveling Wave (BTW) structures (linac 3 and linac 4), with maximum energy gain per cavity of 140 MeV.

The bunch length can be varied by means of two bunch compressors (achromatic magnetic chicanes, BC1 and BC2 in the figure). The charge distribution can be manipulated varying the parameters of an X-band (11.4GHz) linearizer. The trajectory is monitored by 31 stripline Beam Position Monitors (BPMs', [26]) with 20μm rms resolution, and can be modified by changing 24 couples of steering magnets (correctors) both in horizontal and vertical plane. The projected properties of the beam can be monitored by means
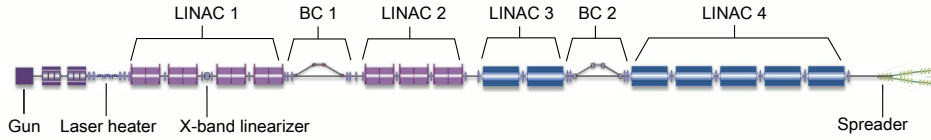
Figure 1: Schematic layout of a the Fermi@Elettra Linac. The machine, composed of 14 accelerating structures, is used to raise the energy of the electrons, provided by the Gun, from 4 MeV to about 1.2 GeV. Linac 1 and linac 2 are composed of SLAC-type accelerating cavities, while linac 3 and linac 4 are composed of BTW structures. There are also two bunch compressors (BC1 and BC2), that are used to control the length of the bunch, and an x-band linearizer. After the acceleration and manipulation, the electrons are transported through the spreader to one of the two FEL beamlines.

of 16 fluorescent multi-screens that can be inserted into the path of the electrons. There are also three rf-deflecting cavities (one after BC1, two at the end of the linac 4) that can be used mainly as diagnostics of the longitudinal phase-space of the electron beam [24, 25]. While the low-energy one was installed before the beginning of the commissioning, only recently one of the high energy ones was installed. The technique we present in this paper was initially developed as an alternative diagnostics for the beam at the linac end.

2.2. Wakefields

Due to both lack of enough BPMs' and correctors around the SLAC-sections of the first part of the machine (Linac 1 and 2), and the presence of the low-energy RF-deflector, which is used as the main diagnostic for the contribution of wakefields in the first part of the linac, the study was carried on the last part of the machine (Linac 3 and 4). This part is composed of BTW accelerating structures, which are characterized by small iris apertures (10 mm) and by geometric transverse wakefields which are stronger if compared to the other structures [27]. As the electron bunches accelerated by the linac are short (less than 10 ps), the transverse wakefields are much stronger than the longitudinal ones [27]. The transverse component of the wakefield plays the major role in the emittance diluition. If we consider a small variation in trajectory with respect to the axis of the accelerating structure, the transverse wakefunction is dominated by the dipole wake[28]. The transverse wakefunction can be approximated with the fol-

4

lowing analytical expression [28]:

$$w(\zeta)[V/C/m^2] = A \left[1 - \left(1 + \sqrt{\frac{\zeta}{\zeta_1}} \right) e^{-\sqrt{\zeta/\zeta_1}} \right] + B\sqrt{\zeta} \ , \qquad (1)$$

where $A = 1.7 \times 10^5$ and $B = 8.5 \times 10^4$ are constants, ζ is the longitudinal coordinate inside the bunch, $\zeta_1 = 1.2 \times 10^{-4} m/l_b$, l_b being the (full width) bunch length. As the transverse wake is dominant, the main effect of the wakefield is that the head of the beam excites tranversal modes which interact with the beam tail, deforming the bunch in the so-called "banana shape". This effect is clearly visible on the multiscreens at different locations of the linac. This is in principle an effect that is dependent on the bunch length. The emittance diluition is due to the transverse deviation of the tail with respect to the head of the bunch. The dipole dominated wake is the key requirement for the method we applied, as the beam dimensions at a screen are directly related to the trajectory and, in first approximation, independent of the optics of the machine. Indeed, a dipole wake does not influence the optical functions of the beam, which means that the same effect can be observed with different optics of the accelerator. On this basis, we believe that the method we propose can be easily extended to other linac-based accelerator machines.

2.3. Feedback

A high degree of beam stability is one of the main requirements for the operations of a linac accelerator facilities. A slow feedback loop [29], based on a Matlab framework, has been developed at FERMI@Elettra and is currently used to steer the beam or restore a given trajectory. In order to satisfy the strictly stability requirements of the Fermi@Elettra project [21], a bunch-by-bunch fast trajectory feedback loop was lately developed. It runs at 50 Hz, the same maximum repetition rate of the machine, and acts bunch-by-bunch, which means that the correction that is calculated for a given bunch is based on the measured position of the previous bunches. In order to maximize the dynamics of the feedback, the correction computation and implementation must be completed within the period between two successive shots, thus 20 ms.

The program reads the beam position at a pre-determined set of BPMs' and sets the steering magnets' strengths in order to correct the trajectory to the desired one. The control algorithm is based on a standard proportional-integral-derivative controller (PID), that can be preceded by different filters, used to cope with spiky data or to remove efficiently periodic noise

sources. In order to correct the trajectory, the feedback needs a response matrix, which correlates the kicks of the correctors and the trajectory distortions. This matrix is calculated by measuring the perturbations in the trajectory produced by each steering magnet: the power supplies of the magnets are driven with a programmed current ramp, while synchronously the BPM positions are acquired. Using the real-time capabilities of the Tango control system, the matrix is measured in about 20 s. After the measurement, the acquired matrix is inverted through Singular Value Decomposition (SVD) algorithm. During this operation correctors, BPMs and singular values can be weighted individually. This further increases the robustness to measurement errors and the correction efficiency, and can also be optimized by means of singular value reduction through the so called Tikonov regularization.

The feedback is routinely used during the commissioning operations and for the first experiments on the FELs' beamlines, to counteract the main noise sources of the machine (slow drifts due to temperature variations, micro interrupts of the beam due to RF plants trips, electric noise, etc.) that could potentially lead to an increase of the trajectory jitter of the machine. The program is also used to restore a given golden trajectory. The measurements presented in this paper took advantage of the advanced capabilities of this trajectory feedback. Preliminary tests were performed using a slow implementation of the feedback. Obtained results are compatible with those obtained using the fast version of the feedback. This observation suggests that the procedure here described can be applied to other machines in which a fast trajectory feedback may not be available.

3. Measurements

In order to evaluate the alignment of the accelerating structures, we performed a series of trajectory scans around one or more accelerating sections. The target position for one or more BPMs' included in the feedback, in fact, can be varied in a continuous way, each different configuration defining a different trajectory bump. The bumps are performed in a controlled way, as the feedback itself closes them and no additional trajectory perturbations are introduced. After the feedback has successfully set the desired trajectory, the transversal profile of the beam is acquired. This is done by looking at the image of the beam, acquired by a Charge Coupler Device (CCD), on one or more fluorescent screens placed downstream the investigated sections. From the transversal beam profile one

can extract directly the information about the horizontal (vertical) beam dimensions, i.e. the standard deviation, $\sigma_{x(y)}$, which in principle would not be enough to characterize the emittance. At first approximation, however, the dominant contribution to the emittance growth along the linac is due to the projected banana shape of the beam. Furthermore we are interested in limited variation of the trajectory around each section, so we think that the beam divergence can be considered constant. As it will be shown in the following, this assumption is confirmed by numerical simulations. All measurements were performed as scans of two variables, an initial BPM, where the trajectory bump was started, and a final BPM where the bump was closed. This kind of 2D scans result in a surface. The usual scan range was of about 3 mm around the nominal center for both BPMs'. In order to determine the relative position between an accelerating section and the upstream and downstream BPMs' only two scans are in principle needed, one for horizontal and one for the vertical plane. The presented measurements were performed with a 350 pC bunch, with bunch length of 2.6 ps (standard deviation) at the gun, compressed to less than 1 ps after the first bunch compressor (compression factor 5). In fig. 2 an example of the obtained surfaces for horizontal (fig. 2(a)) and vertical (fig. 2(b)) alignment is reported. In this case the trajectory bump was performed around the third section of Linac 4. The profile found in both horizontal and vertical planes is related to both the charge distribution in the bunch and the wakefields.

In fig. 2 one can identify a central region where the beam horizontal (vertical) transverse dimension is small, while if the trajectory bump is larger than about half a millimeter the transverse dimension become 50% (or more) larger than the minimum. Both surfaces show high degree symmetry with respect to the diagonals of the scan range. This can be explained a consideration on the dipole wakefields in the section, cfr. fig. 3. If the bump is symmetric (fig. 3(a)), e.g. + 1 mm at entrance, -1 mm at exit, the interaction of the head with the tail in the first half of the accelerating cavity is exactly compensated by one of equal intensity, but opposite sign and there is no net increase of the transverse dimension of the beam. If the bump is parallel (fig. 3(b)), instead, the effect of the interaction between the head and the tail of the bunch builds up along the whole cavity and it is not compensated: the bunch exhibits a pronounced banana shape, and it is easy to understand that the transverse dimension at the screen would be larger. It seems that there is no net gain in choosing a trajectory that lies completely on the axis of the cavity with respect to a symmetric bump. This is true only up to a certain amplitude of the bump, because if the beam passes far from the axis, higher order effects (borders of the cavity, higher

7

order wakefield components) could become important: this effect is barely visible in fig. 2(b), if one notice that the transverse dimension of the beam becomes ~10% larger for symmetric bumps of more than 1 mm.   The
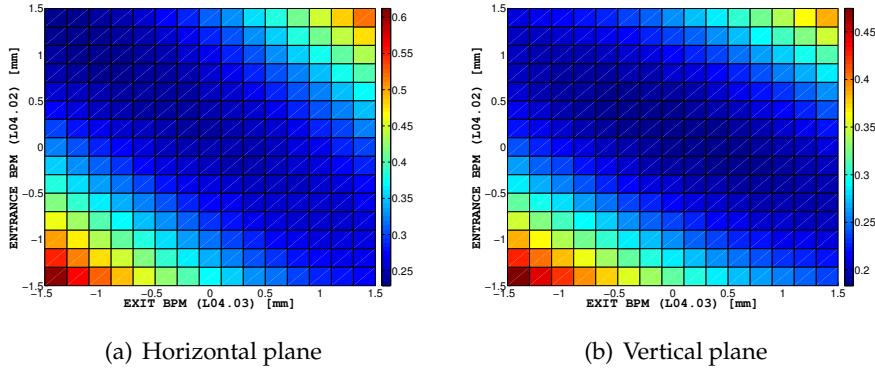

(a) Horizontal plane   (b) Vertical plane

Figure 2: Example of measurements of the transverse σ for horizontal (left panel) and vertical (right panel) planes. The axes of both plots are the entrance and exit positions of the BPMs that were changed during the scan. This two measurements were performed around the third section of Linac 4. The color scale, shown for each figure, is expressed in mm. To obtain these figures, the raw data were cleaned of the spikes and smoothed out.

symmetry of the surface is one of the figure of merit we selected. From the above considerations we can determine that the alignment between the considered accelerating cavity and the upstream and downstream BPMs is, in this case, satisfactory. We can estimate the sensitivity of our measurement to be about 150 μm, which is an order of magnitude larger than the noise in the BPMs' reading, which is of about 20 μm. The reproducibility of the measurement was checked many times and it has always been satisfactory.

3.1. Misaligned cavities

Not all the sections and BPMs were so well aligned as in the example of fig. 2. In case of misalignments or tilts one can observe different shapes of the surface extracted from the transverse profile of the beam. We observed essentially two kind of behaviours, as shown in fig. 4. In presence of a misalignment (fig. 4(a)), we observe that the surface is not symmetric with respect to the symmetric bump axis, but with respect to a shifted axis which is parallel to the symmetric bump one. The misalignment is mainly a rigid translation because the symmetry with respect to the parallel bump axis is

8

TOP VIEW   TOP VIEW

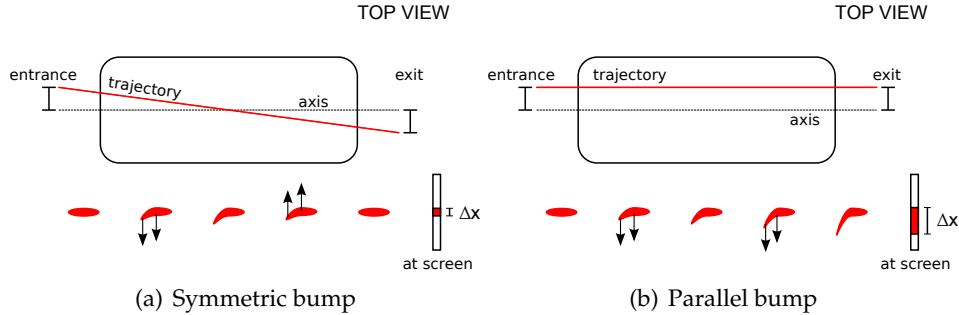(a) Symmetric bump   (b) Parallel bump

Figure 3: Schematic representation (top view) of the different kind of bumps presented in the text. The axis of the accelerating structure and the trajectory of the electrons inside the cavity are shown. On the lower part the effect on the electron bunch is represented, together with the resulting projection Δx at the screen (same considerations can be done for y axis). The arrows represent the effect of the wakefield of the cavity.

good. In this case the position of the minimum can be used as an evaluation of the effective misalignment. In the example, the minimum is at -0.5 mm on both BPMs. In order to correct the misalignment it is sufficient to introduce an offset in both BPMs in order to center the surface at the reference position. Another possibility is the presence of a tilt (fig. 4(b)). In this case the surface is symmetric with respect to an axis, which has an angle with respect to symmetric bump axis. Unfortunately, to cure this situation, a realignment of the accelerating structure and BPMs would be required, which is not always possible.

3.2. Dependence on machine optics

In the preliminary considerations about the wakefield, we underlined that main contribution to the wake is the dipole kick and that the measurement we performed should be independent of the particular setting of the machine optics. Confirming this statement is crucial, because this would allow us to find a golden trajectory that can be used in any of the available machine optics configuration. In fig. 5 we show an example of two scans around the same linac section, performed for different configurations of the machine optics. These were obtained by changing the strength of one quadrupole upstream the cavity itself, by a factor ∼4 from its nominal value. No differences between the two scans are visible, which means that we are able to get the same information about the accelerating section with different optics of the machine. More important, the trajectory we can find with the method can be used with different configurations of the machine.
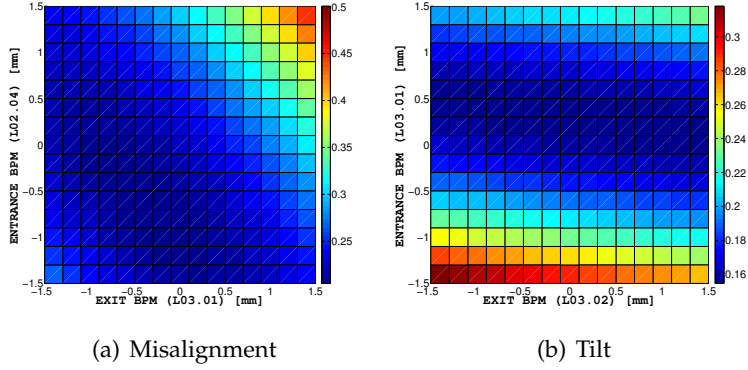
9

(a) Misalignment (b) Tilt

Figure 4: Example of misaligned (left panel) and tilt (right panel) between accelerating cavity and BPM couple. The axes of both plots are the entrance and exit positions of the BPMs involved in the scan. Left panel: measurement performed around the first section of Linac 3; Right panel: measurement performed around the second section of Linac 3. The color scale, shown for each figure, is expressed in mm.
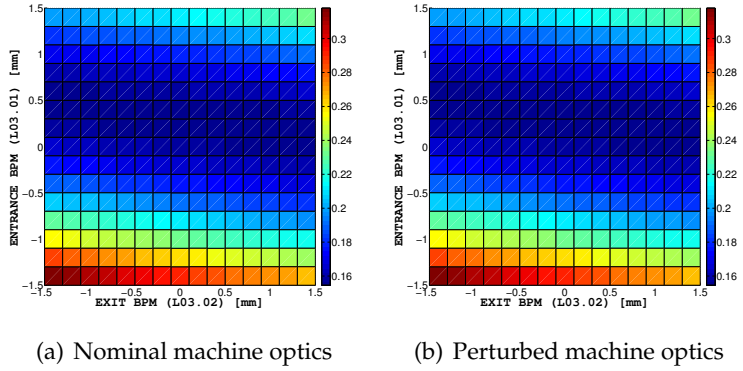

(a) Nominal machine optics (b) Perturbed machine optics

Figure 5: Comparison of two scan around the same accelerating cavity, for two different machine optics configurations. The difference between the optics was obtained by modifying the strength of an upstream quadrupole by a factor ~4. No evident differences in the two scans are present. Measurement were performed around the second section of Linac 3. The color scale, shown for each figure, is expressed in mm.

10

4. Comparison with simulations

In order to better understand the results we have found and to validate the wakefield model for the machine (i.e. eq. (1)), we performed some numerical simulations of the beam dynamics. This is really important for Fermi as the whole matching procedure is based on particle tracking code [31]. The first step consists in trying to reproduce the experimental results. We use elegant [32] simulation code. We implemented in the code the same approach that was used in the measurements, so we performed scans of trajectory bumps and we acquired the beam parameters. In the machine model the wakefields are also taken into account, by means of an external generated input file, which was written using of eq. (1). The simulations have been performed with 100000 macroparticles per bunch and have been carried on the Elettra Scientific Cluster [33].

4.1. Flat beam

As a first step we simulated an electron bunch with flat profile in x, y and z planes. The real beam profile is different from the flat one, but some informations can eventually be extracted. A flat profile is also a good starting point because the computation time required is drastically reduced with respect to a more realistic beam. In fig. 6 a comparison between the measured (fig. 6(a)) and simulated (6(b)) transverse dimension of the beam is shown. The scan is performed around the third section of Linac 4 in both cases, for the vertical plane. The qualitative behaviour is the same for both surfaces, presenting high degree of symmetry with respect to the diagonals of the scanning range. The main difference is that in the simulation the transverse profile grows faster than the measured one, with respect to the parallel bump diagonal. This can be explained in terms of an overestimated wakefield effect on the beam tail. In the real beam, the charge distribution at the tail is lower than the head and, since the effect of the transverse wakefield is proportional to the charge, a larger contribution with respect to the actual beam is reasonable. Simulations can also be useful to check if our hypotheses were reasonable. We are interested in controlling the emittance of the beam, but only the transverse beam size information is accessible, and not the divergence. In fig. 7 a comparison between transverse beam size and emittance is shown. No differences can be found between the two surfaces, which confirms that our hypothesis on the constant beam divergence is reasonable. The values of emittance in fig. 7(b) are a factor 10 less than what is usually measured in the real machine: this is another indication that the real beam is not flat.
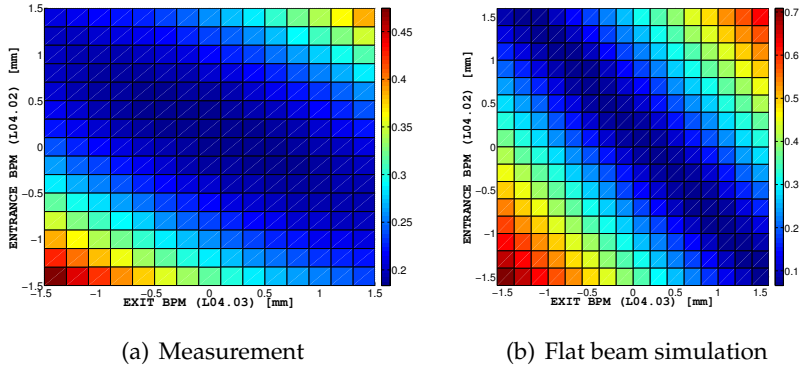
11

(a) Measurement   (b) Flat beam simulation

Figure 6: Comparison between measured (left panel) and simulated (right panel) of the transverse profile of the beam as a function of trajectory bump entrance and exit. Simulated beam profile was flat. The third section of Linac 4 was investigated, and the images are referred to the vertical plane. The qualitative behaviour of the two surfaces is similar. The color scale, shown for each figure, is expressed in mm.
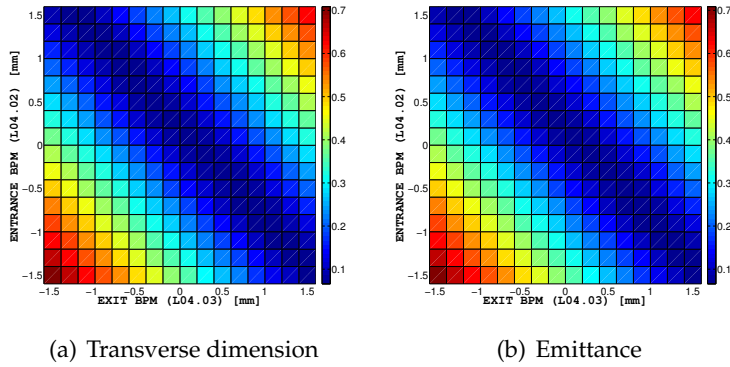

(a) Transverse dimension   (b) Emittance

Figure 7: Comparison between transverse dimension of the beam (left panel) and emittance (right panel) as a function of trajectory bump entrance and exit. The third section of Linac 4 was investigated, and the images are referred to the vertical plane. There is no evident difference in the behaviour of the two surfaces, which confirms our hypothesis on the beam divergence. The color scale is expressed in mm for the beam dimensions and in mm mrad for the emittance.

4.2. Real beam

In order to overcome the limitations we have found for the flat beam case, we performed some simulations using a more realistic beam. This was generated using General Particle Tracer (GPT, [34]) with the same parameters as the electron bunch produced by the gun. In fig. 8 a comparison between the measured transverse dimension of the beam and the simulated one is presented, as a function of the amplitude of the trajectory bumps. The agreement between simulation and measured profile is good, both in qualitative and quantitative terms. The shape of the two surfaces is in fact the same and the rate between maximum and minimum dimensions are comparable. This shows that the model of the machine and the real machine are in good agreement.

(a) Measurement (b) Simulation, realistic beam
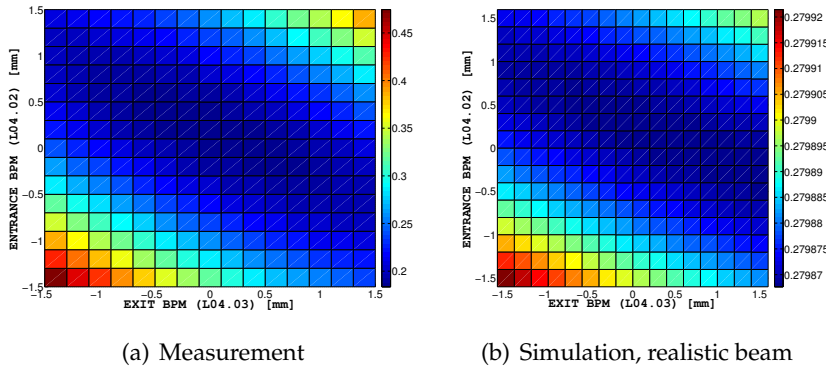
Figure 8: Comparison between measured (left panel) and simulated (right panel) of the transverse profile of the beam as a function of trajectory bump entrance and exit. Simulation was performed using a realistic beam profile. The third section of Linac 4 was investigated, and the images are referred to the vertical plane. The qualitative and quantitative behaviour of the two surfaces is similar. The color scale is expressed in mm.

5. Wakefield model validation

In order to evaluate the agreement between the wakefield model and the measurements we focused on the symmetric diagonal bumps and to the parallel diagonal bumps. In the fig. 9 the simulated and measured values of the transverse profile are reported, as a function of the amplitude of the bump, for the parallel case. The crosses are the measured values, the dashed line is the smoothed curve of the experimental data and the filled lines correspond to two different the simulations, performed using a

13

realistic beam profile and different wakefields: the green line corresponds to a wake as in eq. (1), the light-blue one corresponds to a wake with double the amplitude (factor A in eq. (1)). The agreement between measurements and simulation is good with a wakefunction with double the amplitude.

In order to check independently the agreement between measured data and simulation we showed in fig. 9, we performed a new series of simulations on a different linac section. In fig. 10 the comparison between simulation and measured data is shown, both for horizontal (fig. 10(a)) and vertical (fig. 10(b)) planes. As in the previous case, a wakefunction with double the amplitude lead to a good agreement with the experimental data. The above results show that the method we applied can be useful to evaluate the wakefield contributions for an accelerating structure or other linac elements.
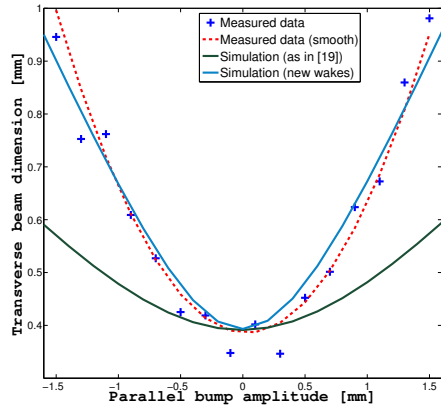
Figure 9: Comparison between measured data (crosses for raw data, dashed line smoothed data) and simulation (filled lines) of the transverse profile of the beam as a function of amplitude of the parallel bump. Simulation was performed using a realistic beam profile and the third section of Linac 4 was investigated. The green line has been obtained using wakefunctions of eq. (1), while light-blue line has been obtained using a wake with double the amplitude. Vertical transverse beam size are reported.

14

(a) Horizontal plane (b) Vertical plane
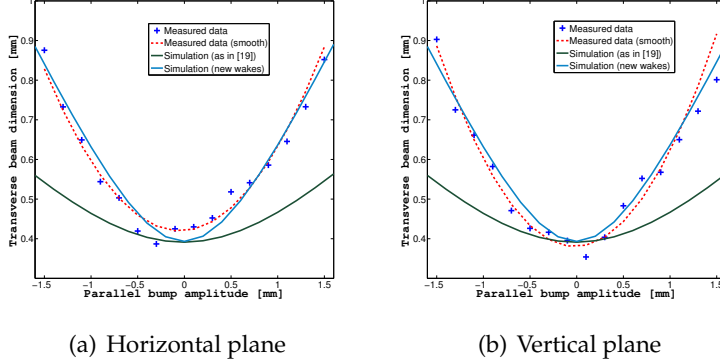
Figure 10: Comparison between measured data (crosses for raw data, dashed line smoothed data) and simulation (filled lines) of the transverse profile of the beam as a function of amplitude of the parallel bump, for horizontal (left panel) and vertical (right panel). Simulation was performed using a realistic beam profile and the fourth section of Linac 4 was investigated. The green line has been obtained using wakefunctions of eq. (1), while cyan line has been obtained using a wake with double the amplitude.

6. Conclusions and prospectives

We presented a beam-based alignment procedure for the accelerating structures of a Linac. It is performed by means of trajectory bump scans and requires the availability of a trajectory feedback. The procedure described above was successfully applied to the last part of the Fermi@Elettra linac (Linac 3 and Linac 4) and took advantage of the capabilities of the Fast Trajectory Feedback here implemented. A figure of merit for the alignment of the accelerating structure has been selected and described and examples of both aligned and misaligned cavities were presented. We demonstrated that the procedure does not depend on a particular configuration for the machine optics. The alignment of all the 7 BTW sections has been performed and an "optimal" trajectory was found. The net effect that was measured is a reduction of the emittance at the linac end by a factor $\sim$2 in the horizontal (from 6 mm mrad to 3 mm mrad) and a $\sim$30% (from 2.6 mm mrad to 1.8mm mrad) improvement in the vertical plane. This is a significant improvement of the beam quality, and can positively impact on the FEL performance.

The measured profiles were also compared to particle tracking simulations. We determined that the main contribution to the emittance dilution of the banana shape can be evaluated by measuring the transverse profile of the beam only. This confirms the hypothesis of negligible contribution

15

of the beam divergence to the emittance during this kind scans. We found a satisfactory agreement between measurements and simulations when a realistic beam profile was used. The wakefield model of the machine has been tested with measurements on the machine. Further investigation will be necessary in order to investigate the discrepancy in amplitude of the wakefunction we have found. We are planning to investigate the possible usage of this method as a bunch length diagnostics. Further, we would like to evaluate the effect of the trajectory bumps and emittance growth directly on the FEL performance. The procedure here described is completely general and in our opinion can be successfully applied to other linac-based facilities.

7. Acknowledgements

References

[1] C. W. Roberson and P. Sprangle, A review of freeelectron lasers, *Phys. Fluids B* **1**, 3 (1989).

[2] J. Rossbach, A VUV free electron laser at the TESLA test facility at DESY, *Nucl. Instrum. Methods Phys. Res. A* **375** 269 (1996).

[3] `http://xfel.desy.de/`

[4] TESLA - Technical Design Report, DESY 2001-011, ECFA 2001-209, TESLA Report 2001-23, TESLA-FEL 2001-05, Deutsches Elektronen Synchrotron, Hamburg, Germany (2001).

[5] K. Vinit et al., Electron beam requirements for Smith-Purcell backward wave oscillator with external focusing, *Phys. Rev. STAB*, **12**, 070703 (2009).

[6] C. Montag, Electron Beam Stability Requirements for Linac-Ring Electron-Ion Colliders, *Proceedings of PAC 2005, Knoxville, Tennessee*, (2005).

[7] F. Iandola, Electron Beam Focusing for the International Linear Collider, *American Association for the Advancement of Science Annual Meeting, Boston, MA* (2008).

[8] http://marie.lanl.gov/docs/thp163.pdf

[9] J. Buon, Beam Phase Space and Emittance, *CAS - CERN Accelerator School : 5th General Accelerator Physics Course*, Jyväskylä, Finland, (1992).

[10] M. Sands, The Physics of Electron Storage Rings. An Introduction, SLAC, (1970).

[11] A. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, *Whiley, New York*, (1993).

[12] K. L. F. Bane, Wakefields of Sub-Picosecond Electron Bunches, *SLAC-PUB-11829*, (2006).

[13] K. L. F. Bane, M. Sands, Wakefields of Very Short Bunches in an Accelerating Cavity, *SLAC-PUB-4441*, (1987).

[14] A. Chao et al., *SLAC-PUB-2498*, (1980).

[15] R. Wanzenberg, Review of Beam Dynamics and Instabilities in Linear Colliders, *Proceedings of LINAC 1996, Geneva, Switzerland*, (1996).

[16] R.W. Assmann, Beam Dynamics in SLC, *Proceedings of PAC 1997, Vancouver, Canada*, (1997).

[17] P. Tenenbaum, Effectiveness of Emittance Bumps on the NLC and US Cold LC Main Linear Accelerators, *SLAC-TN-04-038*, (2004).

[18] J. Seeman, The Introduction of Trajectory Oscillations to Reduce Emittance Growth in the SLC Linac, *SLAC-PUB-5705*, (1992).

[19] C. Adolphsen et al., Flat Beam Studies in the SLC Linac, *Proceedings of PAC 1993, Washington DC, USA*, (1993).

[20] E. Prat, Dispersion effects on performance of the free-electron laser FLASH, *Phys. Rev. STAB*, **13**, 040701 (2010).

[21] C.J. Bocchetta, et al., *FERMI@elettra CDR*, http://www.elettra.trieste.it/FERMI/index.php?n=Main.CDRdocument, (2007).

[22] W. Colson, The Nonlinear Wave Equation for Higher Harmonics in Free-Electron Lasers, *IEEE J. Quantum Electron.*, **17**, 1417 (1981).

[23] R. Bonifacio et al., Physics of the High-Gain FEL and Superradiance, *Riv. Nuovo Cimento*, **17**, 9 (1990).

[24] M. Petronio, PhD Thesis, (2010).

[25] Michael Röhrs, PhD Thesis, (2008).

[26] Libera Single-Pass detector, manufactured by Instrumentation Technologies.

[27] P. Craievich et al., The Short-Range Wakefields in the BTW Accelerating Structure of the ELETTRA Linac, *Technical Note STM-0402*, Sincrotrone Trieste, (2004).

[28] P. Craievich et al., Single-bunch emittance preservation in the presence of trajectory jitter for FERMI@elettra-seeded FEL, *Nucl. Instrum. Methods Phys. Res. A* **604** 457 (2009).

[29] M. Lonza et al., Beam-Based feedbacks for FERMI@Ekettra Free Electron Laser, *Proceedings of IPAC'10, Kyoto, Japan*, (2010).

[30] G. Gaio et al., Commissioning of the FERMI@Elettra Fast Trajectory Feedback, *Proceedings of ICALEPCS'11, Grenoble, France*, (2011).

[31] S. Di Mitri et al., Integration of Elegant Tracking Code into the Tango Server-Based High Level Software of FERMI@Elettra for Optics Measurements and Modeling, *Proceedings of IPAC'10, Kyoto, Japan*, (2010).

[32] M. Borland, elegant: A Flexible SDDS-Compliant Code for Accelerator Simulation, Advanced Photon Source LS-287, September 2000.

[33] `http://gan.elettra.trieste.it/`

[34] `http://www.pulsar.nl/`

Origin of Shift Dependent Multipoles in Apple-II Undulators.

Mirko Kokole

Kyma Tehnologija d.o.o., Kraška ulica 2, 6210 Sežana

Abstract

APPLE-II insertion devices are very flexible devices for production of variably polarized photons. This devices inherently suffer from shift dependent integrated field multipoles that can reach values which can seriously deteriorate quality of the electron beam. Since there is no really effective shimming method for correction of this errors, it is important to understand where they originate. Paper presents a study of integrated field multipoles shift dependency based on deformation of magnetic array due to magnetic forces. We have modeled separately deformations of each magnet keeper in the magnetic array. Model calculations have shown that most of integrated field multipole dependency on the shift is due to mechanical deformation in combination with magnetic effects.

Keywords: undulators, eliptically polarized undulator (EPU), integrated field multipoles, skew quadrupole, advanced planar polarized light emitter II (APPLE-II)

1. Introduction

Advanced Planar Polarized Light Emitter II (APPLE-II) [1] is a variable polarization insertion device able to produce all possible photon polarization modes linear, elliptical and inclined linear. These devices have been used in most of the worlds 3rd generation light sources. APPLE-II device is composed of four girders of magnet arrays as show in the figure 3. Two diagonal girders are usually movable and two are fixed, though there are devices with all four girders movable. The central part of magnetic array of composed of Halbach array [2]. The end field terminations are different for each device. The longitudinal displacement of movable girders if called shift or phase. At zero shift the magnetic field polarization is vertical

therefore giving horizontal polarization of light. The half period shift will produce vertical polarization. Shifts in between will give different elliptical photon polarizations. Energy variation is achieved by opening and closing the gap between upper and lower girder.

Insertion device should have minimal influence of the electron beam in all modes of operation, consequently requiring a tight control of the integrated magnetic field over the length of the device. Both second and first field integrals (see eq. 1) should be as close to zero as possible in all operational modes. Also the field integrals mutipoles (see eq. 2) should be as low as possible and should change as little between different operational modes. For this reason a lot of design effort was put in the magnetic structure, specifically on the end field terminations [3]. But all this efforts do not solve all the possible errors. It has been show in multiple cases [4, 5, 6] that mechanical deformations play a critical role in the performance of this devices. Most notable notably shift dependent multipoles have been observed in the EPUs for FEL-1 and FEL-2 line at FERMIatElettra [7, 8] and also at other light sources [4]. For these reasons a study of the origin of shift dependent multipoles was prompted.

2. Integrated Field Multipoles

Integrated field multipoles are derived from the transverse change in the first field integral (see eq. 1). They are defined as the polynomial expansion around certain transverse point, usually on the optical axis. Eq. 2 defines normal (vertical) A_n and skew (transverse, horizontal) B_n integrated field multipoles. For dipole is $n = 0$, quadrupole $n = 1$ and sextupoles $n = 3$. They all have different effects on the electron beam. Dipole gives a angular deviation, quadrupole produces beam focusing and sextupole effects the betatron tune. In and ideal case they should all be zero or as close to zero as possible [9].

$$I_x(x,z) = \int_{-L/2}^{L/2} B_x(x,z,s)ds$$

$$I_z(x,z) = \int_{-L/2}^{L/2} B_z(x,z,s)ds$$

$$I2_x(x,z) = \int_{-L/2}^{L/2} sB_x(x,z,s)ds$$

$$I2_z(x,z) = \int_{-L/2}^{L/2} sB_z(x,z,s)ds$$

(1)

2

Even in ideal case it is not possible to keep all the multipoles to zero at all operational modes of an APPLE-II insertion device. Normal dipole is generated in the symmetric field device by default and changes with the shift due to the non zero permeability of magnetic material. Sextupole and its change is also generated by the non zero permeability of the magnetic material. It is possible to limit this changes by proper design of the end field terminations [3].

$$I_x(x, z_0) = \sum_{n=0}^{\infty} A_n(z_0) x^n$$

$$I_z(x, z_0) = \sum_{n=0}^{\infty} B_n(z_0) x^n$$

(2)

On the other hand it is not possible to generate the quadrupoles both normal and skew in ideal device. Hence it is believed that they arise for the mechanical deformations. For example a rotation of the horizontally magnetized block around vertical axis [6] can give a rise to a skew quadrupole, which is the most notably observed shift dependent multipole [4, 7, 8]. In the 55 mm periods EPUs for FEL-1 for FERMI@Elettra a large shift dependent skew quadrupole was observed. Largest change of about 400 G was observed at minimum 10 mm gap figure 1. Values of multipoles decrease with increasing gap.
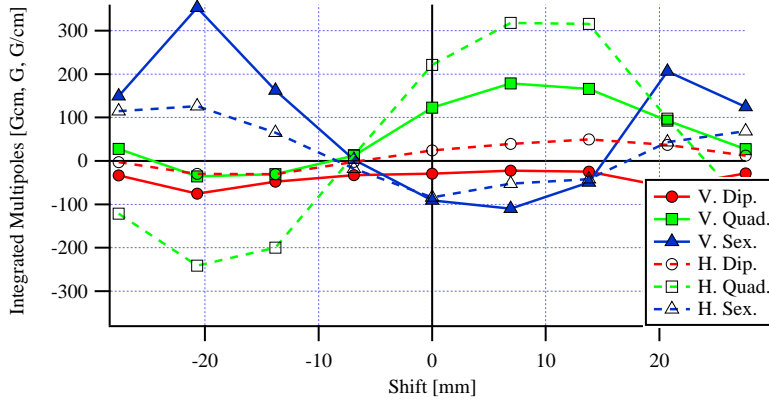
Figure 1: Measured change of integrated field multipoles with shift. 55 mm period EPU for FERMI@ELETTRA at 10 mm gap.

To understand which deformations will generate different multipoles and how they change with shift, it is necessary to make a model that includes

3

also deformations for the ideal structure. Following paragraphs contain description and results from such a model.

3. Magnetic model with structural deformations due to magnetic forces

In order to study how mechanical deformation influence the magnetic performance of the insertion device a special RADIA code was developed without any symmetries allowing a movement and rotation of each magnet block. This code generates a APPLE-II magnetic structure and applies deformations to the structure by displacement and rotation of each block. It is possible to either enter deformations from external files, produced by the mechanical Finite Element Analysis (FEA) analysis of the mechanical structure or the code can also calculate simplified deformations due to magnetic forces.

3.1. Magnetic forces

First part of the code will calculate magnetic forces and torques on each magnet block, hence giving an input for the FEA mechanical calculations. For this calculation permeability of material was not taken into account since it does not effect the forces significantly. Largest force is exerted in vertical direction for both horizontally and vertically magnetized block. This force is mostly due to the neighbors in the array. Horizontal and transverse force change with the shift due to the changing neighbors.
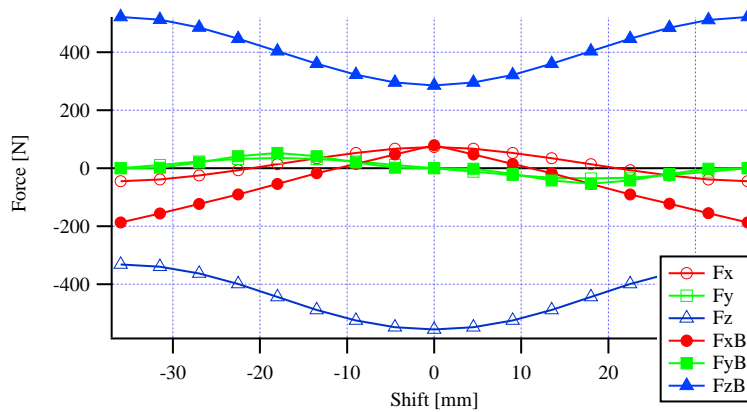
Figure 2: Calculated change of forces with the shift in a vertically and horizontally polarized magnetic block.

4

At the moment no available computer is able to compute a complete mechanical structure applying forces on each magnet separately. So it is necessary to apply some assumptions. Hence a series of steps was studied in order to get a complete picture. By applying only the deformations from frame and girders no change in multipoles is observed. From this it can be concluded that the multipoles must have a origin in different displacement of each magnet block. Since FEA can not calculate deformation in a complete mechanical structure for each block separately a approximation approach was taken. Magnetic forces and torques on each magnet were calculated and from this an isotropic deformation, translation and rotation, was calculated by simple multiplication of deformation coefficient. Figure 3 show a small model with deformed structure, deformations are exaggerated 100x. For each configuration of gap/shift a new calculation of forces is required, therefore a very long calculations are required to achieve a clear picture.
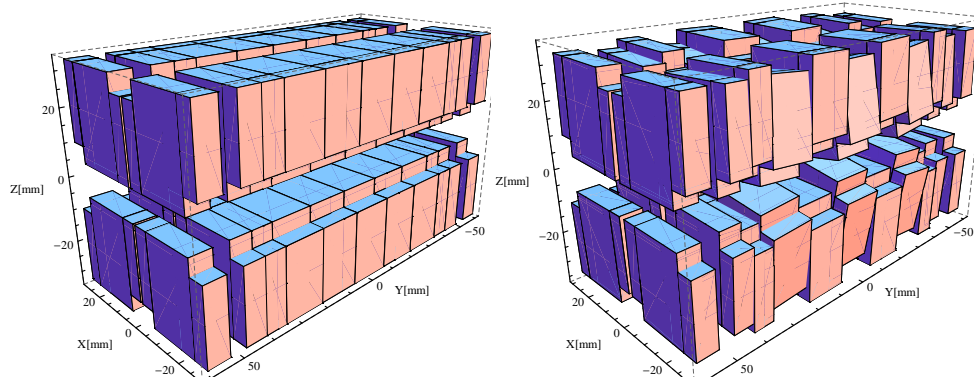
Figure 3: Short one period 3D model of APPLE-II insertion device (left). There are four girder, two on lower beam and two on upper beam. Usually two diagonal girders are movable. The displacement of movable girders in longitudinal Y axis is called a shift. 3D model of a deformed APPLE-II magnetic structure (right). Deformation are amplified 100 x for clarity.

3.2. Non deformed case

First let us examine a non deformed case. For a model parameters for FERMI EPUs are taken (see table 1) . Figure 4 shows the transverse change of normal and skew field integrals are minimum gap and shift of 10 mm. As it can be seen there is no change in the skew field integral. If we look at the

shift dependency of integrated multipoles we see that only vertical dipole and sextupole change with the shift. This is due to non unit permeability of magnetic material. We must note here that it is not possible to generate quadrupoles in a non deformed APPLE-II magnetic structure.

Parameter	Value
Period	55 mm
Gap	10 mm
No. of Periods	6
Block width	28 mm
Block height	28 mm
Br	1.25 T
(χ_{par}, χ_{per})	(0.05, 0.15)

Table 1: Parameters for a model magnetic structure.
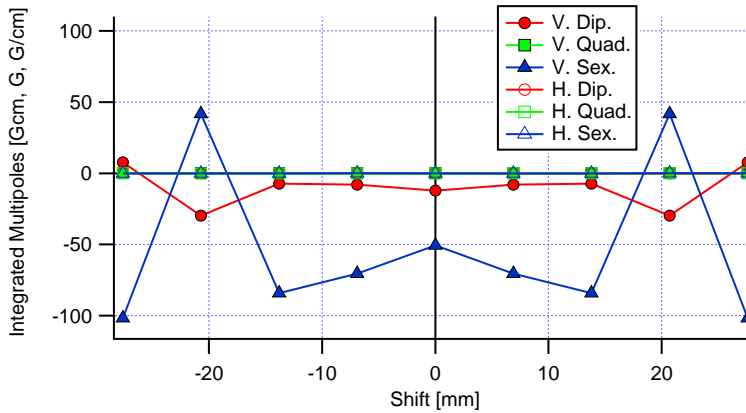
Figure 4: Calculated shift dependency of integrated multipoles for non deformed magnetic structure. Dipole and sextupole changes are due to non zero magnetic permeability of magnetic blocks. Quadrupoles can not be generated in such a structure.

To understand whether multipoles are generated in the end field termination or are cumulative through the magnetic structure we need to look at the change of multipoles with increasing number of periods. Only very small changes in the integrated multipoles is observed so we can conclude that integrated multipoles are generated in the end field termination and are not cumulative trough the magnetic structure. Here we keep in mind that this is only valid for our model where deformations are isotropic and only due to magnetic forces.

6
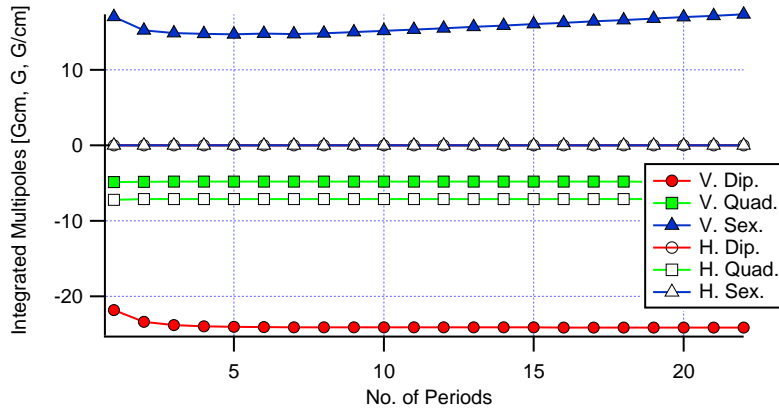
Figure 5: Calculated change of integrated multipoles with number of periods. It is evident that multipoles are mostly generated in the end field section and in close proximity to them.

3.3. Deformed case

A detailed study of influences of displacement in different directions and rotations around different axes has shown that only rotation round longitudinal axis can generate a significant skew quadrupole. On the other hand sextupole, specifically normal sextupole can be generated almost in all cases of translations and rotations. From this we can conclude that most important contribution to quadrupole generation is from rotation around longitudinal axis. Again this only applies for isotropic deformation from magnetic forces only and when deformations are applied to all magnetic blocks in a structure. As noted before also rotation around vertical axis of the longitudinally magnetized block generate a skew quadrupole behavior, but this is only true if we do not rotate all the blocks in the magnetic structure.

Attempts to reproduce the shift dependency of skew quadrupole as measured in the FERMI undulators have failed, as it is apparently not possible to reproduce such an error by applying translations and rotations according to magnetic forces and torques on single magnets only. Magnet displacements (translations and rotations) along different degrees of freedom will not occur with the same deformation constants, since they are held by a holder that breaks the symmetry. For a reproducible description of the shift dependency a detailed finite element analysis (FEA) of magnet deformations and their application to the entire magnetic structure is needed.

However measurements of an EPU for Pohang Light Source II (PLS2), which

7
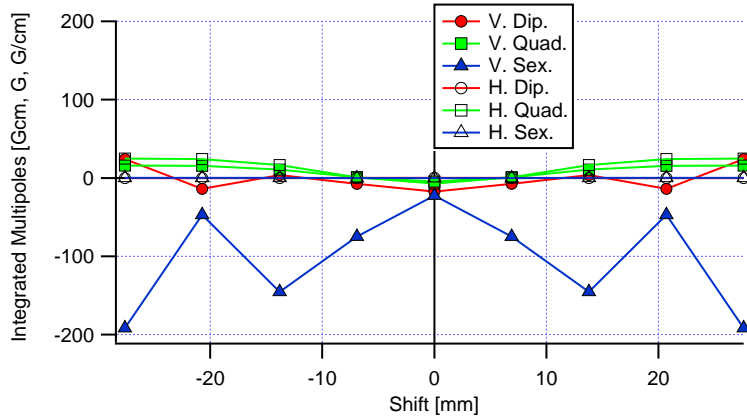
Figure 6: Calculated shift dependency of integrated multipoles on a deformed model. Quadrupoles are generated by rotation around longitudinal axis.

has different improved mechanical design of a magnet holder, give much better accordance with a model. Figure 7 shows comparison of a change of integrated skew quadrupole with the shift change for a model magnetic structure and measurements of a 70 mm period EPU for PLS2. We can see that in this case he model describes well the behavior of integrated skew quadrupole.
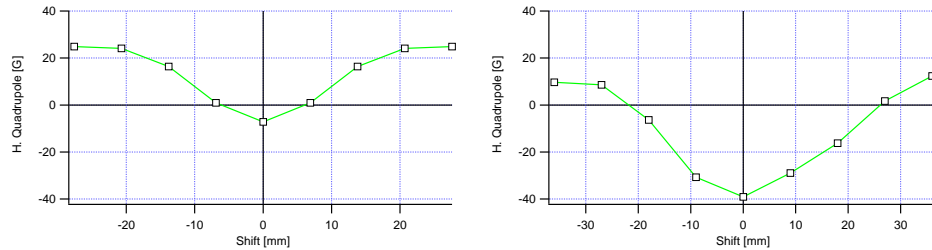
Figure 7: Comparion of shift dependency of skew quadrupole in a measured (right) 70 mm period EPU and model calculation (left). We see that we can at least qualitatively predict the behavior of quadrupoles. Calculation is only valid for and isotropic deformations due to magnetic forces.

4. Conclusions

We have presented a model of APPLE-II device that includes deformation of magnetic structure due to magnetic forces. As it can be seen in some cases where deformations can be described as isotropic and only due

to magnetic forces, we get good results when comparing simulation and real measurements. But when deformations are not isotropic we need to do real FEA modeling of compete magnetic structure applying appropriate forces on each magnet block. At the moment such calculation is beyond capability of a normal personal computer and we have to apply some simplification. Such a model with anisotropic deformations is currently under development.

References

[1] S. Sasaki, K. Miyata, T. Takada, A new undulator for generating variably polarized radiation, Japanese Journal of Applied Physics 31 (Part 2, No. 12B), (1992), L1794–L1796.

[2] G. Brown, K. Halbach, J. Harris, H. Winick, Wiggler and undulator magnets, a review, Nuclear Instruments and Methods in Physics Research 208, (1983), 65–77.

[3] J. Chavanne, P. Elleaume, P. V. Vaerenbergh, End field structures for linear/helical insertion devices, Proceedings of the 1999 Particle Accelerator Conference, New York, USA.

[4] C. Steier, et al., Study of row phase depenedent skew quadrupole fields in apple-II type epus at the ALS, Proceedings of EPAC04, Lucerne, Switzerlan.

[5] F. Briquez, et al., Construction of short period apple ii type undulators at soleil, Proceedings of FEL'08, Gyeongju, Korea.

[6] C. Kitegi, F. Briquez, M. E. Couprie, T. E. Ajjouri, J. M. Filhol, K. Tavakoli, J. Veteran, Dvelopment of a short period high field apple-ii undulator at soleil, Proceedings of IPAC'10, Kyoto, Japan.

[7] M. Kokole, B. Diviacco, T. Milharcic, G. Soregaroli, M. Tedeschi, M. Zambelli, Magnetic characterization of the FEL-1 undulators for the fermi@elettra free-electron laer, Proceedings of FEL2010, Malmo, Sweden, 2010.

[8] M. Kokole, B. Diviacco, T. Milharcic, , M. Zambelli, Magnetic characterization of the FEL-2 undulatros for the fermi@elettra free-electron laser, Proceedings of the IPAC 2011, San Sebastian, Spain, 2011.

[9] Y. C. Chae, G. Decker, Advanced photon source insertion device field quality and multipole error specification, http://www.jacow.org.

Meshless numerical solution of the diffusion equation with Dirichlet and Neumann boundary conditions

Katarina Mramor

Laboratory for Advanced Materials Systems, The Centre of Excellence for Biosensors, Instrumentation and Process Control - COBIK, Velika pot 22, SI-5250 Solkan, Slovenia

Abstract

The diffusion equation with Dirichlet and Neumann boundary conditions is solved for a two-dimensional cylinder. Both analytical and numerical solutions are obtained in order to test the accuracy of the meshless method. In this article, the local radial basis function collocation method (LRBFCM) is presented and applied to a two-dimensional diffusion equation. The results obtained are tested for a variety of different time steps and time intervals as well as a choice of different grids. These sample calculations show that the numerical scheme provides results that are in good agreement with analytical solutions.

Keywords: meshless method, diffusion equation, analytical solution, numerical solution, cylindrical coordinates

1. Introduction

The diffusion equation is a partial differential equation which describes the physical processes such as heat conduction, spread of dye in a stationary fluid, population dispersion, and other similar processes. In this paper, the focus is on heat transfer and thus the equation predicts the variation of temperature in a given region over time. Heat conduction equation can be derived from the principle of energy conservation. The diffusion equation is a standard problem in thermodynamics which can be solved with different numerical methods [1–5] such as finite elements method, finite volume method, finite differences method or meshless method. A novel approach to solve the heat conduction equation by a meshless method is presented in this paper.

Email address: `katarina.mramor@cobik.si` (Katarina Mramor)

The most notable advantages of the Local Radial Basis Functions Colloca-
tion Method (LRBFCM), which is the meshless method used in this case,
are the independence of the method on the dimensions of the problem, the
ability to solve the governing equations of the problem in their strong formu-
lation, and the ability to perform calculations on random nodes, that need
not be geometrically connected. A case study is made for a two-dimensional
diffusion equation in cylindrical coordinates with axial symmetry, compar-
ing analytical results with those obtained with a meshless numerical method.
The aim of this paper is to test the accuracy of the solutions obtained by a
LRBFCM.

2. Model

The diffusion equation in a most general way is stated as [6]:

$$\frac{\partial}{\partial t}C(\vec{r}, t) = D\nabla^2 C(\vec{r}, t), \tag{1}$$

where C represents the sought for variable at position $\vec{r} = (r, \phi, z)$ and time
t, and D represents diffusion constant. When C represents temperature T
the equation is also called heat conduction equation and D then represents
thermal diffusion. As the equation is used to solve heat conduction in a
cylinder, it is here rewritten in cylindrical coordinates:

$$\frac{\partial C}{\partial t} = D\left(\frac{1}{r}\frac{\partial C}{\partial r} + \frac{\partial^2 C}{\partial r^2} + \frac{\partial^2 C}{\partial z^2} + \frac{1}{r^2}\frac{\partial^2 C}{\partial \phi^2}\right). \tag{2}$$

Due to the symmetry of azimuthal coordinate, Equation 2 is reduced to two
dimensions:

$$\frac{\partial C}{\partial t} = D\left(\frac{1}{r}\frac{\partial C}{\partial r} + \frac{\partial^2 C}{\partial r^2} + \frac{\partial^2 C}{\partial z^2}\right). \tag{3}$$

2.1. Initial and boundary conditions

The calculations are made for a two-dimensional cylinder, with maximum
radius r_0 and maximum height h. The calculation domain scheme with
boundary conditions and initial conditions is shown in Figure 1. The initial
conditions are stated as:

$$C(r, z, 0) = \begin{cases} 0K; & r, z \in \Gamma, \\ 1K; & r, z \in \Omega, \end{cases} \tag{4}$$

where Γ represents the edge of the calculation domain and Ω its interior.
Dirichlet boundary condition $T_{BC} = 0$ K is applied to the top, bottom

and right side of the domain, whereas the Neumann boundary condition: $\frac{\partial T}{\partial \bar{n}}|_{r=0} = 0$ is applied to the left side of the domain, which represents the center of the cylinder.
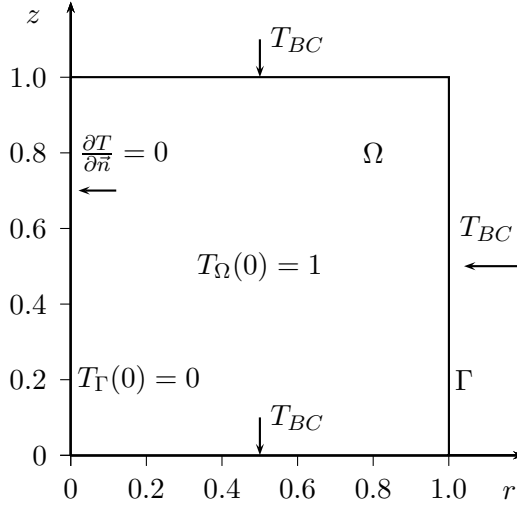

Figure 1: Domain scheme of two-dimensional cylinder with initial and boundary conditions. The initial temperature inside the domain is 1 K, whereas the initial temperature on the domain boundaries is 0 K. Dirichlet boundary condition is applied to the top, right and bottom side of the domain and Neuman boundary condition is applied to the left side of the domain.

2.2. Analytical solution

In order to solve the diffusion equation analytically, a technique called the separation of variables is used. In this technique, a solution is acquired by separating the time- and space-dependent variables:

$$C(\vec{r},t) = R(r)\Phi(\phi)Z(z)\tau(t), \qquad (5)$$

where $\tau(t)$ is space independent function, while $R(r)$, $\Phi(\phi)$ and $Z(z)$ are time independent functions; each dependent only on one spatial variable. By substituting the separable solution (Equation 5) into Equation 2, gathering the time-independent terms on one side and dividing them by Equation 5 the following equation is derived:

$$\frac{1}{D\tau}\frac{\partial \tau}{\partial t} = \frac{1}{rR}\frac{\partial R}{\partial r} + \frac{1}{R}\frac{\partial^2 R}{\partial r^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} + \frac{1}{r^2\Phi}\frac{\partial^2 \Phi}{\partial \phi^2} = -k^2. \qquad (6)$$

3

Since the left side of Equation 6 is a function of time only and the right side of the same equation is the function of space only, both sides must be equal to a constant $-k^2$. The following equations are thus obtained:

$$\tau \sim e^{-Dtk^2}, \tag{7}$$

$$Z \sim e^{\pm ik_p z}, \tag{8}$$

$$\phi \sim e^{\pm im\phi}, \tag{9}$$

$$R \sim J_m(k_{mn}r), \tag{10}$$

where J_m stands for a Bessel function, and coefficients k_{mn} and k_p are expressed as $k_{mn} = \frac{\zeta_{mn}}{r_0}$, where ζ_{mn} is zero of the Bessel function, and $k_p = \frac{p\pi}{h}$.

Taking into account the symmetry of ϕ axis, it follows that $m = 0$. The model 'loses' a dimension (in azimuthal direction) and is thus reduced into a two-dimensional model. A solution can therefore be written as:

$$C_{mn}(r, z, t) = A_{mnp} \, J_m(k_{mn}r) \, \sin(k_p z) \, e^{-Dk^2 t}, \tag{11}$$

where A_{mnp} is a constant. In general, the sum of solutions

$$C(r, z, t) = \sum_n A_{0np} \, J_0(k_{0n}r) \, \sin(k_p z) \, e^{-Dk^2 t} \tag{12}$$

that satisfies the boundary conditions must also satisfy the initial conditions. By using an initial condition $C(r, z, 0) = 1$ we obtain

$$A_{0np} = \frac{8}{p \, \pi \, R \, k_{0n} \, J_1(k_{0n}R)}, \tag{13}$$

where J_1 is Bessel function of the first kind. Figure 2 shows the analytical solution of the heat equation for two-dimensional cylinder. The temperature field is depicted for times 0.1 s, 1.0 s and 2.0 s thus giving the estimation of the temperature diffusion time frame. The pictures on the left side (Figures 2a, 2c and 2e) show the temperature values in degrees K for the whole two-dimensional calculation domain at elapsed times 0.1, 1 and 2 s. As expected, the cooling starts at the outer walls (top, bottom and right wall) and with time moves towards the center of the cylinder. The highest temperature, consequently at all times remains in the central point of the left side wall. The radial dependence of temperature (T(r)) which is depicted in the right side pictures (Figures 2b, 2d and 2f) confirms the results as they clearly show, that the highest temperatures are located at the center of the cylinder and that the lowest temperatures are located at the cylinder outer walls.
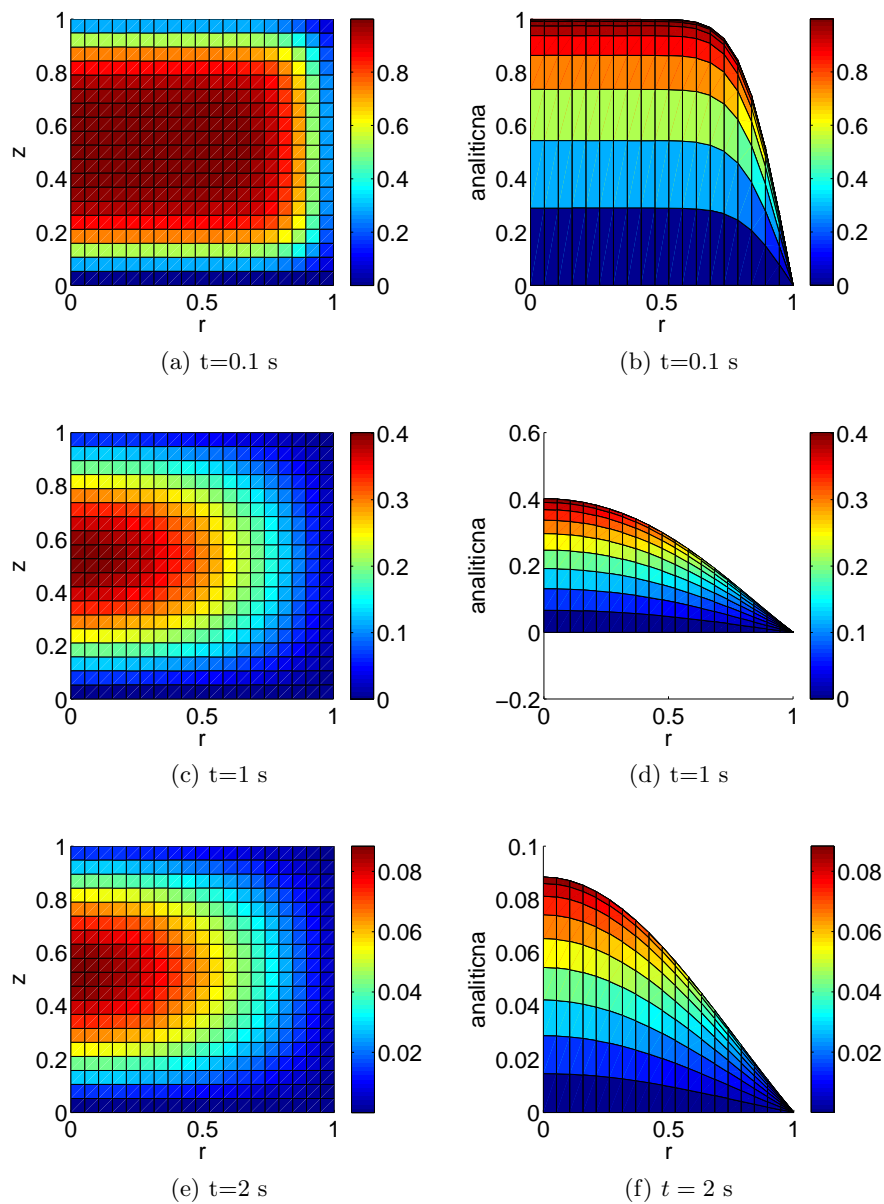
Figure 2: The evolution of temperature field calculated for two-dimensional cylinder presented for three distinct times: $t = 0.1$ s, $t = 1$ s and $t = 2$ s. The left pictures represent the top view of temperature field T(z,r) with different colours presenting different values of temperature whereas the right graphs depict the dependence of temperature on the radius of the cylinder T(r).

3. Computational methods

in general, the temperature field is a continuous function of space and time and must therefore be described by an infinite number of degrees of freedom. The problem is reduced in to the solution of algebraic equations by replacing the continuous function by formulae with finite degrees of freedom. For our case, the meshless or mesh reduction technique is applied for spatial discretization and the finite differences method (FDM) is used for time discretization.

3.1. Meshless method for spatial disretization

Mesh reduction technique is a numerical simulation algorithm that uses a set of arbitrary distributed nodes both on the boundary and within the computation domain to represent the solution of a physical phenomena. The simplest among the variety of meshless methods developed to solve heat and fluid flow equations is the Radial Basis Function Collocation Method (RBFCM). In this paper, its local version will be implemented. The idea behind Local Radial Basis Function Collocation Method (LRBFCM) [7], [8] is to approximate the function locally over a set of neighbouring nodes using radial basis functions (RBFs) as a basis and the collocation to find the coefficients. In this case multiquadric (MQ) RBF forms Ψ are used:

$$\Psi(\vec{r}) = \sqrt{\frac{r^2}{c^2} + 1}, \tag{14}$$

where r is the distance between the reference point and its neighbours and c is the predetermined shape parameter. Equation 3 is replaced with its discrete approximation

$$\nabla^2 T(\vec{r}, t) \approx \sum_{j=1}^{N} \Psi_n \, \Psi_{nj} \, T_j(\vec{r}, t), \tag{15}$$

where T stands for temperature, t for time, $\vec{r}$ is the position of reference point, Ψ_{nj} are components of the inverse of the basis matrix and Ψ_n are components of the basis matrix on which the operator, in this case ∇^2, is applied. The boundary values are computed as $T(\vec{r}, t) = T_{BC}$ for Dirichlet boundary conditions and as:

$$\frac{\partial T}{\partial \vec{n}} = 0 = \left(\frac{\partial \Psi_n}{\partial n_r} + \frac{\partial \Psi_n}{\partial n_z} \right) \Psi_{nj} T_j \tag{16}$$

for Neumann boundary condition, where $\vec{n}$ is normal to the boundary, with n_r and n_z representing its components.

3.2. Finite difference method for time discretization and final discretisized equation

FDM [9] is an approximation technique in which a solution is approximated by replacing the derivatives with difference quotients:

$$\frac{\partial T}{\partial t} \approx \frac{T(\vec{r}, t_{k+1}) - T(\vec{r}, t_k)}{\Delta t}, \tag{17}$$

where Δt is the time difference between time t_{k+1} and t_k.

By applying the spatial and time discretization to the diffusion equation

$$\frac{\partial T}{\partial t} = D \ \nabla^2 T, \tag{18}$$

the equation thus becomes

$$T(r_i, t_{k+1}) = T(r_i, t_k) + \Delta t \ D \ \Psi_n \ \Psi_{nj} T_j(r_i, t_k). \tag{19}$$

3.3. Error analysis

The meshless method is tested by solving the heat equation (Equation 3) and by comparing the obtained solutions to those calculated analytically (Section 2.2). The absolute error ε is calculated as the absolute value of the analytically T_{an} and numerically T_{num} calculated temperatures:

$$\varepsilon = |T_{num} - T_{an}|. \tag{20}$$

Additionally, the mean absolute error MAE is calculated as an average of the absolute errors:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |T_{num} - T_{an}|, \tag{21}$$

where N stands for the number of points in which the temperatures are calculated.

The propagation of error is first researched for a variety of different time steps Δt and numbers of time steps N_t. The elapsed time t from the beginning of the observation is obtained as a product of time step and the number of time steps:

$$t = \Delta t \ N_t. \tag{22}$$

4. Results

The results were calculated for two-dimensional cylindrical domains with 25, 100 and 400 points. The number of points in azimuthal (z) and radial (r) direction is equal; that means that the domain with 25 points has 5 points in radial and 5 points in azimuthal direction. The points are evenly distributed. The results are obtained after 1 s of elapsed time, unless otherwise stated. The analysis of time steps, time intervals and final elapsed times are done on the domain with 10 points in each direction. In all the examined cases, N_r stands for the reference number of the point, with 1 suiting the radius $r = 0$ m and 10 suiting the radius $r = r_0$. The results of the analysis give the discrepancy between numerical and analytical values at the central (5th) point in azimuthal direction (z). To show the accuracy of the results ε and MAE are calculated for each of the proposed examples.

The error dependency on the number of calculation points is tested on domains with 5, 10 and 20 nodes in each direction. It is expected that the more accurate results are obtained for domains that have more points. The reliance between the error and the number of nodes in the mesh is tested for a $N_t = 10^6$ time steps of the size $\Delta t = 10^{-6}$ s; that is after 1 s from the start of observations. The actual temperature profiles T(z,r) for analytical (left: Figures 3a, 3d and 3f) and numerical (right: Figures 3b, 3c and 3e) solutions of a given problem for geometry with 25, 100 and 400 calculation points are shown in Figure 3. Even though, the domain with the highest number of points gives the best resolution of temperature profile, the domain with moderate number of points, which is in our case 100, is chosen for further calculations; specifically for the dependency of the method on the time step and time interval. The moderate sized domain is chosen due to the satisfactory values of both ε and MAE that are achieved for the domain with 100 points and slightly smaller calculation times. Even though, the calculation times are not significant in our case, they might become a problem in other cases with more complex computational domains and governing equations. The difference between both solutions, analytical and numerical, is best observed from Figure 5, where a row of central points is depicted for each calculated case. Calculated MAE values are presented in Figure 4. As predicted, the absolute error ε as well as the mean absolute error MAE (Figure 5) both lessen as the number of points in the domain increases. The error reaches the largest value at the center of the cylinder ($N_r = 1$) and diminishes towards its boundary ($N_r = 10$).

(a) 25 calculation points

(b) 25 calculation points

(c) 100 calculation points

(d) 100 calculation points

(e) 400 calculation points
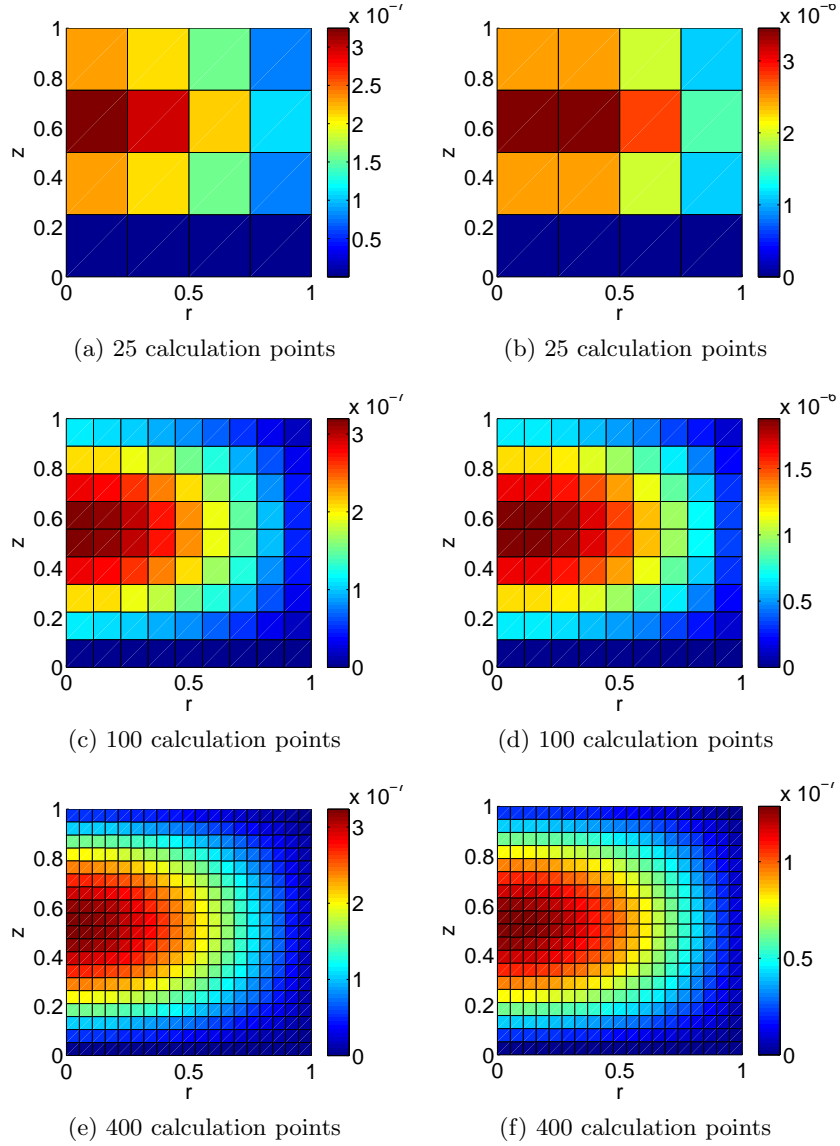
(f) 400 calculation points

Figure 3: Analytical and numerical results of the heat equation for two-dimensional cylinder calculated for the calculation domains with 25, 100 and 400 calculation points. The left graphs represent analytical results whereas the right graphs represent numerical solutions.
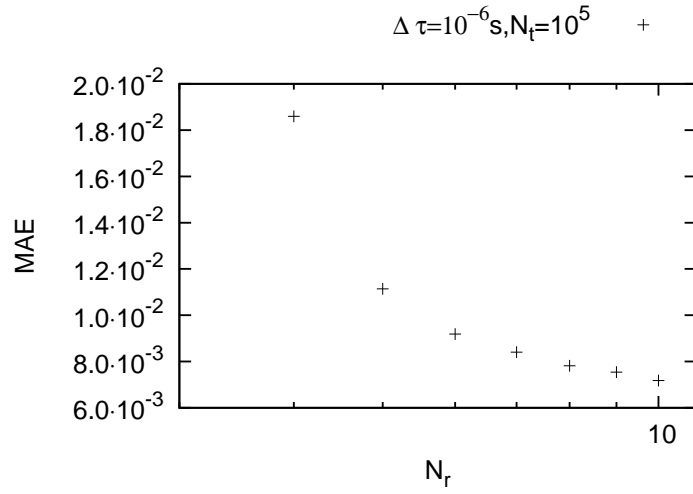
Figure 4: Mean absolute error (MAE) dependent on the number of nodes in a mesh for a time step of $\Delta t = 10^{-6}$ s and the number of time steps $N_t = 10^5$. The number of points in the domain for the elapsed time of 0.1 s are 16, 25, 36, 49, 64, 81 and 100. As the number of points in the domain increases the MAE decreases.
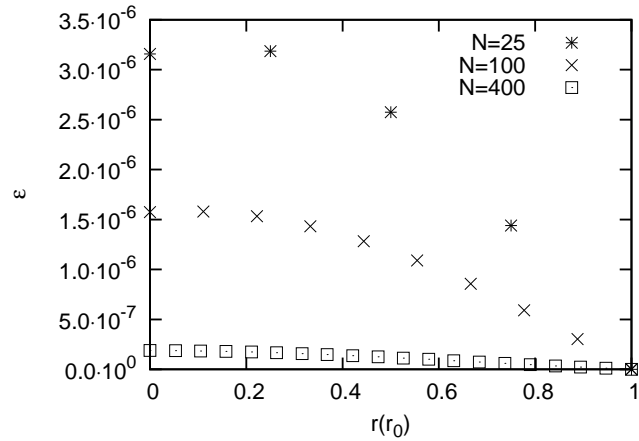
Figure 5: Absolute error ε dependent on the number of nodes in a mesh for a time step of $\Delta t = 10^{-6}$ s and the number of time steps $N_t = 10^6$. The number of points in the domain for the elapsed time of 1 s are 25, 100 and 400. For each case ε is depicted for $z = 0.5$ h.

The analysis of time steps is done on a domain with 100 calculation points and the results are obtained after 0.1 s and 1 s of elapsed time. The MAE has been calculated for all points of the domain for every case and is shown in Figure 6 for 0.1 s of elapsed time. From the results in the figure it can be deduced that MAE decreases by increasing the number of time steps and decreasing the size of the time steps for a constant number of nodes. The ε was calculated for 1 s of elapsed time and is for time steps smaller or equal than $\Delta t = 10^{-4}$ s similar (Figure 7), whereas the error gained in the case with shorter time step is slightly smaller. Error trends are similar in all the cases; reaching the largest values at the center of the cylinder ($N_r = 1$) and diminishing towards its boundary ($N_r = 10$).
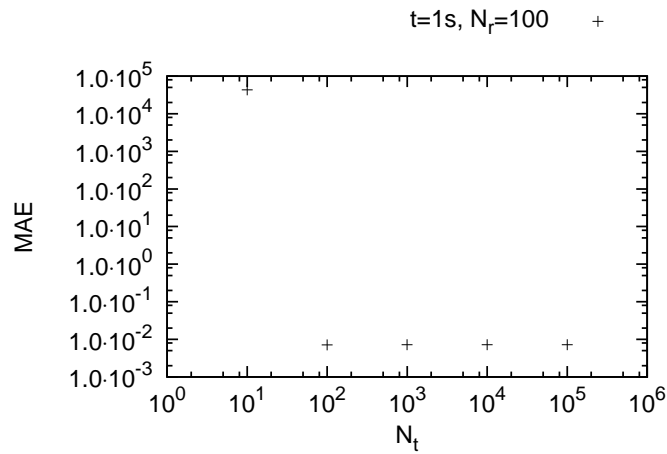
Figure 6: Mean absolute error (MAE) for the elapsed time $t = 0.1$ s, with different time steps Δt and time intervals N_t for domain with 100 calculation points. For a constant elapsed time MAE diminishes when the size of the time step is decreased and the number of time steps is increased. The MAE converges towards a constant value.
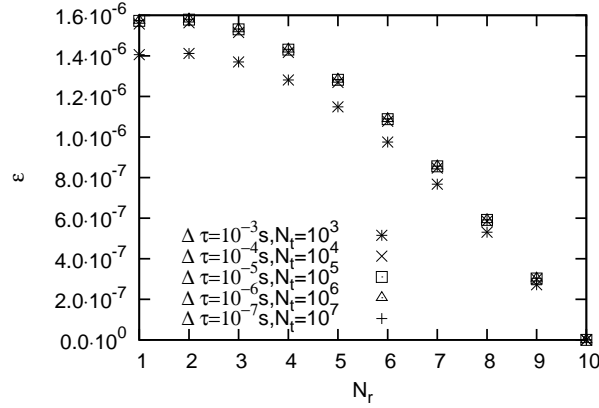
Figure 7: Absolute error ε dependence on the elapsed time t, for different time steps Δt and time intervals N_t for domain with 100 calculation points. Error trends reach the largest values at the center of the cylinder ($N_r = 1$) and the smallest at its edge ($N_r = 10$).

The dependency of the error on the number of time steps is checked for a domain with 100 points. The time step is held constant at $\Delta t = 10^{-6}$s and the number of time steps is varied from 10^2 to 10^6. Consequently, the elapsed time t ranges from 10^{-4} to 1 s. MAE diminishes with time, which means that with longer observation times we achieve more accurate solutions, as the temperature relaxes toward steady state. MAE are graphically represented in Figure 8. The lowest ε (Figure 9) is obtained for the longest elapsed time. The trends for $t < 1$ s are similar and show a dip at $N_r = 4$, which is due to the temperature gradients. As the temperature relaxes towards steady state, the temperature gradients become smaller, thus resulting in an error curve without a dip.
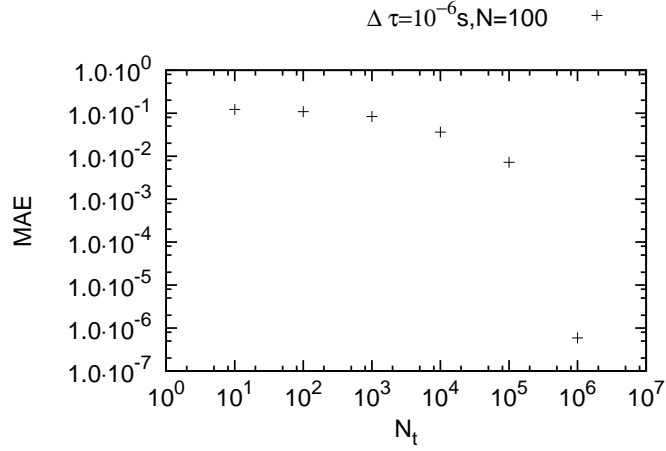
Figure 8: Mean absolute error (MAE) dependency on the number of time steps N_t for 10 x 10 domain for $\Delta t = 10^{-6}$ s. The elapsed time t ranges from 10^{-4} s to 1 s. MAE diminishes with time as the temperature relaxes towards steady state. This means that the results are more accurate for longer observation times.
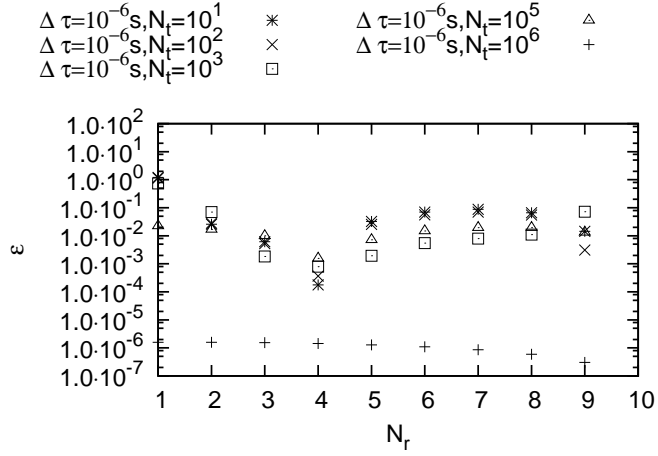
Figure 9: Absolute error ε dependency on the number of time steps N_t for 10 x 10 domain for $\Delta t = 10^{-6}$ s. The elapsed time t ranges from 10^{-4} s to 1 s. A dip at $N_r = 4$ in temperature trends is due to the temperature gradients. As expected, the lowest ε is obtained for the longest elapsed time. This curve does not have a dip, as the temperature gradients diminish as the temperature reaches steady state.

13

5. Summary and conclusions

In this paper, the analytical solution of two-dimensional diffusion equation in cylindrical coordinates is presented along with numerical solutions of the same equation. It is shown that the LRBFCM scheme which is used to obtain the numerical results provides solutions that are in good agreement with analytical results. The obtained calculations coincide with predictions made for denser meshes, larger time steps and longer time intervals. Since the method proved to be in good agreement with expected results, LRBFCM can be applied to more complex cases of computational fluid dynamics.

References

[1] P. Johns, A simple explicit and unconditionally stable numerical routine for the solution of the diffusion equation, International Journal for Numerical Methods in Engineering 11, (1977) 1307–1328.

[2] A. Kienle, M. Patterson, Improved solutions of the steady-state and the time-resolved diffusion equations for reflectance from a semi-infinite turbid medium, Journal of the Optical Society of America A: Optics, Image Science, and Vision 14, (1997) 246–254.

[3] D. Anderson, J. Tannehill, R. Pletcher, Computational fluid mechanics and heat transfer, Hemisphere Publishing, New York, NY, (1984).

[4] N. Ozisik, Finite difference methods in heat transfer, CRC Press, Boca Raton, (1994).

[5] P. Siegel, R. Mosé, P. Ackerer, J. Jaffré, Solution of the advection–diffusion equation using a combination of discontinuous and mixed finite elements, International journal for numerical methods in fluids 24, (1997) 595–613.

[6] I. Kuščer, A. N. Kodre, H. Razpet, M., B. Golli, Matematika v fiziki in tehniki, DMFA, Ljubljana, (1994).

[7] B. Šarler, R. Vertnik, Meshfree explicit local radial basis function collocation method for diffusion problems, Computers & Mathematics with Applications 51, (2006) 1269–1282.

[8] R. Vertnik, B. Šarler, Meshless local radial basis function collocation method for convective-diffusive solid-liquid phase change problems, International Journal of Numerical Methods for Heat & Fluid Flow 16, (2006) 617–640.

[9] G. Smith, Numerical solution of partial differential equations: finite difference methods, Oxford University Press, USA, (1985).

Application of Artificial Neural Networks in Design of Steel Production

Tadej Kodelja

Laboratory for Advanced Materials Systems, The Centre of Excellence for Biosensors, Instrumentation and Process Control, Solkan, Slovenia

Abstract

Artificial neural networks (ANNs) have been developed in this work as an alternative to physical modelling for computing the final mechanical properties of steel semi products. Production of steel bars for application in forging, spring and engineering industries in Štore Steel company has been used as a demonstration of this new methodology. The complete process chain consist of six processes: melting of steel, continuous casting of steel, hydrogen removal, reheating, rolling and finally cooling on a cooling bed. Two open source ANN libraries have been used (Aforge and NeuroDotNet). Both libraries contain neural computing elements that have the ability to respond to input stimuli and to learn to adapt to the environment. The process chain is completely defined by 154 process parameters. The approximation model was built on basis on 34 process parameters that turn to be influential and also vary over the data used. Five output values were observed: elongation (A), tensile strength (Rm), flow limit (RP), hardness after rolling (HB) and shrinkage (Z). The results obtained in parametric studies based on the ANN based model seem consistent with expectations based on industrial experiences. However further improvements in data acquisition and analytical procedures are envisaged in order to obtain reliable enough methodology for use in everyday industrial practice.

Keywords:
neural networks, response approximation, material processing, steel manufacture

Email address: `tadej.kodelja@gmail.com` (Tadej Kodelja)

1. Introduction

For several years artificial neural networks have been successfully used for second level process automation in basic industries [1]. For example, in steel making industries neural networks are already being used for predicting steel mechanical properties after heat treatment [2], for thermal model of a ladle furnace [3], for rolling mills [4], for heat transfer in continuous casting process [5]. One of the fields where it is possible to exploit neural networks is to predict five important mechanical properties of steel (elongation, tensile strength, flow limit, hardness and shrinkage) on the basis of their composition and other process parameters that define the complete processing path.

The complete steel manufacturing process in the Store Steel company [6] is schematically represented in Figure 1. There are six individual processes [7–9]: steel making, continuous casting of steel, hydrogen removal, reheating, multiple stage rolling, and cooling on the cooling bed. Each of this process can be modelled either by a physics based numerical model or by an artificial neural network model. Output values of a process sometimes define the next process in the chain and thus act as input parameters (e.g. defining initial or boundary conditions) in the model of that process. Another possibility when using ANN is to build a common model of the whole production chain. In this case, we can model only outcomes after the last process and relate them to process parameters that can vary in the system. The aim of this work is to explore the possibility of applying artificial neural network for the whole process chain. The result of the ANN model will be used for prediction of five mechanical properties for steel manufacturing process chain in terms of process parameters.

2. Steel manufacture process and process parameters

The manufacture of steel involves a series of processes. First, the steel is melted from scrap in an electric arc furnace (EAF). EAF heats charged material by means of an electric arc. When steel temperature is around 1600 degrees Celsius the molten steel is then poured into a preheated ladle - this is called tapping. During tapping process we add lime, manganese, silicon, carbon and, if necessary according to the type of steel that is being made, other alloying elements. The ladle is then transported by a crane to the refining station, where more carbon, manganese, silicon and vanadium are added to achieve the specification of the grade of steel being made. Argon gas bubbles through the ladle to help remove any remaining impurities.
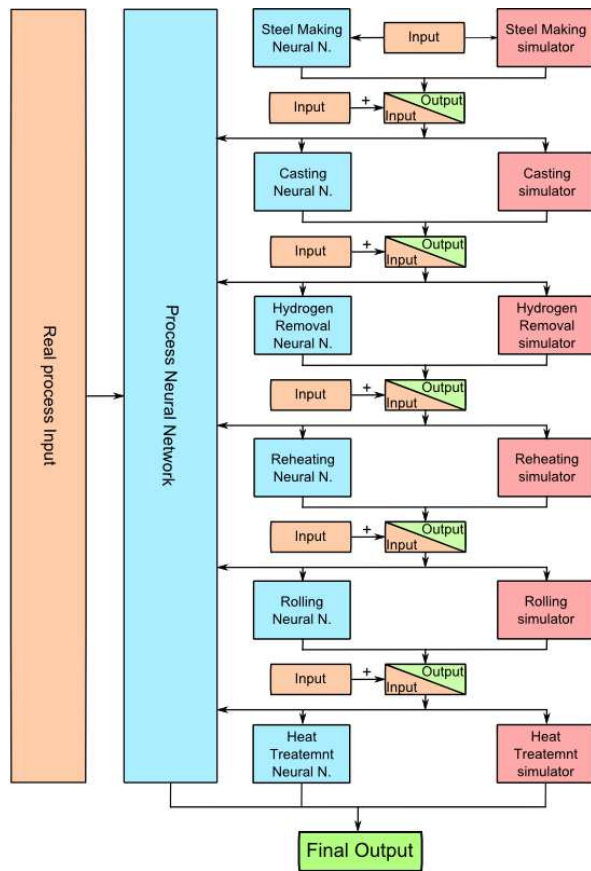
Figure 1: Complete steel manufacture modelling scheme is showing how steel process path can be modelled. Each of six processes can be modelled separately or one master model for the whole process path can be done.

Once the specification of the steel is confirmed the ladle is transported by the crane to the continuous casting area. Here the ladle gate is opened and the molten steel is allowed to flow in a controlled matter into a tundish. This steel flows into three billet moulds. The steel billets are formed and then cut to the desired length for further processing [10, 11]. Once the appropriate billet grade is determined the billet must be reheated and is loaded into gas reheat furnace. The furnace is continually loaded, so as one billet enters the furnace another one that is now fully heated is rolled out ready for the rolling process. The reheating furnace typically takes two to three hours to heat up to the rolling temperature of 1260 degrees Celsius. Gas burners in the furnace provide the heat and hot exhaust gas is reused to preheat the

incoming combustion air, minimizing the heat loss. The billets are moved through the furnace on a walking beam which lifts the entire furnace full of billets and moves them forward one step at a time. Once heated to the appropriate temperature, the billet enters the continuous bar rolling line. In basic terms the rolling line has two phases: the "roughing" stands or preliminary rolls, which do the initial rough shaping, and then the "finishing" stands which finally shape the steel to the appropriate size, deformation pattern and shape. The steel can pass through as many as eighteen sets of rolls to reach its final shape. As the steel billet passes through each roll it speeds up as the area of the bar gets smaller, the length gets longer and therefore it accelerates. Each of the rolls is technically designed to not only produce the correct size product, but do it in a way that ensures the product quality, and considers the limitations of the way steel behaves, and of the equipment that controls and drives the rolls. After the steel leaves the rolling line it goes to the cooling bed. The bar on the cooling bed is cut to length with the flying shear as it enters the bed. This ensures each bar can fit on the bed. Once on the cooling bed the bar cools to a manageable state and is then further processed at a cold shear where it is cut to the customer required length. The product is then bundled and labelled to provide its unique identity. Complete process is shown in Figure 2.
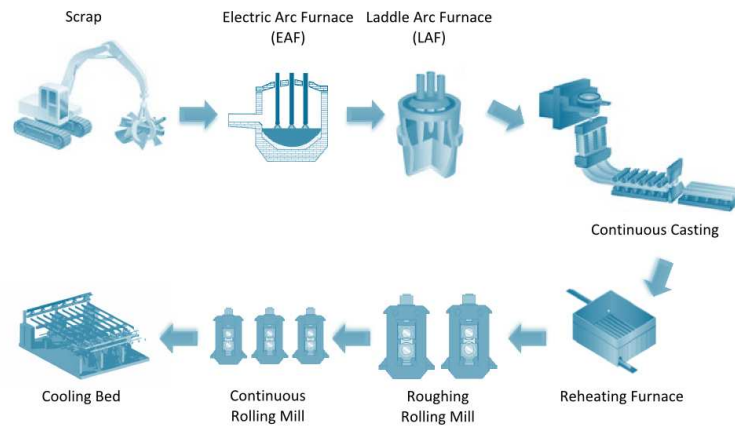
Figure 2: Steel manufacture process path in Store Steel company is composed of several processes which are schematically shown in the picture above.

There are 123 important parameters divided into in seven groups that define complete process chain (Table 1). There are 24 parameters defining the steel grade, 12 parameters for casting, 2 parameters for hydrogen removal, 4 parameters for reheating furnace, 31 parameters for rolling mill, 43 parameters

for continuous rolling mill and 7 parameters regarding cooling bed. On the other hand we have five basic mechanical properties of the materials that represent output values (Table 2).

Table 1: Complete list of process chain (input) parameters.

ID	PROCESS	PARAMETER	USED
1-24	Compositions	Elements: C, Si, Mn, P, S, Cr, Mo, Ni, Al, Cu, Ti, V, W, Sn, As, Zr, Ca, Sb, B, N, O, H, Pb, Zn	24
25	Continuous casting of steel	Casting dimensions (140 x 140mm or 180 x 180mm)	
26		Casting temperature	1
27		Casting speed	1
28		Casting powder type	0
29		Mould level depth	0
30		Mould water flow	1
31		Mould inlet water temperature	1
32		Mould outlet water temperature	
33		Wreath spray flow	1
34		Wreath spray temperature	0
35		Spray cooling system 01 spray flow	1
36		Cooling water 01 temperature	1
37	Hydrogen removal	Time in the furnace	
38		Temperature in the furnace	0
39	Biller reheating furnace	Conveyor speed	
40-42		Temperature in furnace Zone $1-3$	3
43	Rolling mill	Input dimension (140 x 140mm or 180 x 180mm)	
44		Input temperature	
45		Number of rolling passes	
46-52		Entry rolling speed pass $1-7$	
53-59		Radius of roll $1-7$	
60-66		Roll gap $1-7$	
67-73		Roll groove $1-7$	

74	Continuous rolling mill	Input dimension (140 x 140mm or 180 x 180mm)	
75		Input temperature	
76		Entry or outlet rolling speed	
77-86		Roll 1 – 10 engagement yes/no	
87-96		Radius of roll 1 – 10	
97-106		Roll gap 1 – 10	
107-116		Roll groove 1 – 10	
117	Cooling bed	Product dimension – cross-section	
118		Product dimension – length	
119		Product temperature	
120		Distance between two products	
121		Number of bars in one spot	
122		Lifting apron (radiation shield) height	
123		Frequency of product moving	
123	**TOTAL**		**134**

Table 2: Process chain (output) values.

ID	TYPE	VALUES	USED
1	Mechanical properties of materials	elongation (A)	1
2		tensile strength (Rm)	1
3		flow limit (RP)	1
4		hardness after rolling (HB)	1
5		and shrinkage (Z)	1
5	**TOTAL**		**5**

3. Neural Networks Approximation Module

In several practical cases the process design parameters must be adapted quickly in order to produce results that comply with customer requests. With classical approach to optimization of process parameters, long computational times needed for each run of the process simulation at trial design parameters can therefore limit applicability of process optimization in industrial environment. Solution has been conceived in the form of approximation

6

of system response [12], which is calculated on basis of sampled response prepared in advance either by runs of numerical model or by measurements performed on previous designs used. The optimization procedure that produces process design parameters consistent with the current requirements is then performed on the surrogate model based on the approximated response.

In the current work, approximation of process chain based on neural networks is considered. A convenient characteristic of neural networks is that approximation is performed in two separate stages (Figure 3). In the training stage, the network is trained by using the sampled response (either measured or calculated by a numerical model). In the approximation stage, trained network is used for all subsequent calculations of approximated response at arbitrary values of input parameters.
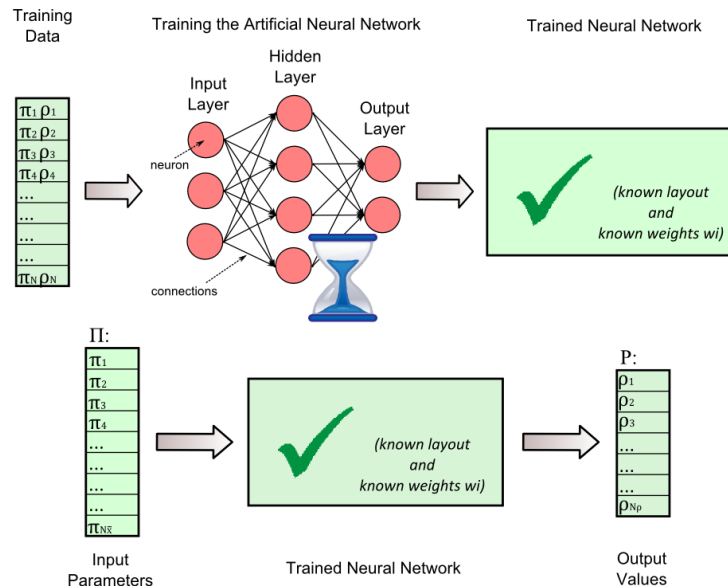

Figure 3: Approximation with neural networks: training a network with presented data (top) and calculation of approximated response with trained network (bottom).

An in-house approximation module has been built based on general purpose neural network libraries. It features modular design such that new underlying libraries can be easily utilized (open source libraries Aforge.Net [13] and NeuronDotNet [14] are currently used). This also provides good flexibility in integration with other software, designing training strategies, filtering training data, verification of results, testing different network layouts, etc. This is crucial when approximating behaviour of material processing sys-

tems with large number of processing parameters. Data obtained from such systems is often inaccurate or even corrupted due to practical limitations in acquisition procedures. Response sampling can not be planned in advance but is accommodated to production schedules in the factory, therefore information available may be deficient in some regions of parameter space in order to obtain good response approximation and therefore verification of results plays an important role.

4. Training The Artificial Neural Network

We trained artificial neural network with data from the complete steel production chain in Store Steel Company. Process is completely defined with 123 process parameters (Table 1). 34 influential input parameters and 5 output values were taken into account. Process data for steel bars for applications in the forging, spring and engineering industries was used. After separating data belonging to two billet dimensions (140mm and 180mm) and after some filtering to exclude corrupted data, 1879 data sets for dimension 140mm have been prepared. The data have been manually collected from different synchronized data bases. The main goal of the study is to train the artificial neural network in order to be capable of predicting elongation, tensile strength, flow limit, hardness after rolling and shrinkage, while changing the chemical composition and other process parameters accounted for in training procedure. For practical set-up of the relevant artificial neural network we used Aforge.Net [13] and NeuroDotNet [14] libraries, implemented in our software module written in C-sharp. Datasets were stored in predefined JSON-based format and imported from file before training. Our module allows us to check training and verification errors during the training procedure. The procedure consist of five steps: reading data from a file, data preparation, training, testing and prediction of unknown output values based on different combinations of 34 input parameters listed in Table 1. During training, state of the artificial neural network is adjusted to data sets with known output values. These comprise historical cases of steel production in the past. During training, the ANN response in training and verification points is checked in order to see how well it does at predicting known and unknown output values. Verification and training points used for testing are usually a subset of historical data. Verification points are randomly chosen from datasets before training starts and are not used in training procedure, while training points are. When error on training points becomes smaller than the user specified tolerance or when the number of training cycles reaches specified number, the training stops. We tried

8

different combinations of layouts and training parameters. We made more than 20 trainings with both, NeuronDotNet and Aforge libraries. The best results were achieved by using NeuronDotNet library, one hidden layer with 15 neurons. The learning rate was set to 0.3, momentum to 0.6 and maximum number of epochs was set to 100000. Training procedures were always performed on HP workstation HPDL380G7 with 12 Intel Xenon 2.0GHz processors, 24GB installed RAM. Trained neural network which gave us the best results was trained in approximately 18 hours.

5. Parametric Studies

After training of the neural network was done, errors of the approximated outputs in verification sets were calculated and some parametric studies were performed. With these parametric tests we try to determine the accuracy of the ANN and also verify dependences between parameters.
In the first study we check relative errors of the obtained approximation in all verification points. These errors are a good indicator of accuracy of the obtained neural network-based approximation and are defined as follows:

$$\delta_{vi} = \left| \frac{v_{real}(\vec{p_i}) - v_{approx}(\vec{p_i})}{v_{range}(\vec{p_i})} \right|, \tag{1}$$

where $v_{real}(\vec{p_i})$ is the actual value of output quantity v in the verification point No. i, $v_{approx}(\vec{p_i})$ is the approximated value of this quantity in the same point in parametric space, and $v_{range}(\vec{p_i})$ is the range of the considered quantity over all training sets. Division by $v_{range}(\vec{p_i})$ is performed for normalization and easier comparison of results for different quantities that may typically differ by several orders of magnitude. $\vec{p_i}$ is the vector of parameters corresponding to the verification point i, in which actual values of output quantities are known, since verification points are taken out of the provided data.
Verification points in our test represented five percent of the complete data-set provided. There were 1879 points with corresponding values of output quantities in the data-set, of which 94 were randomly selected as verification sets and were excluded from training procedure. This preparation procedure is done automatically before training starts. Actual and predicted values of elongation, tensile strength, flow limit, hardness after rolling and shrinkage are shown in Figure 4. Maximum relative errors in verification points for elongation, tensile strength, flow limit, hardness after rolling and shrinkage are shown in Figure 5.
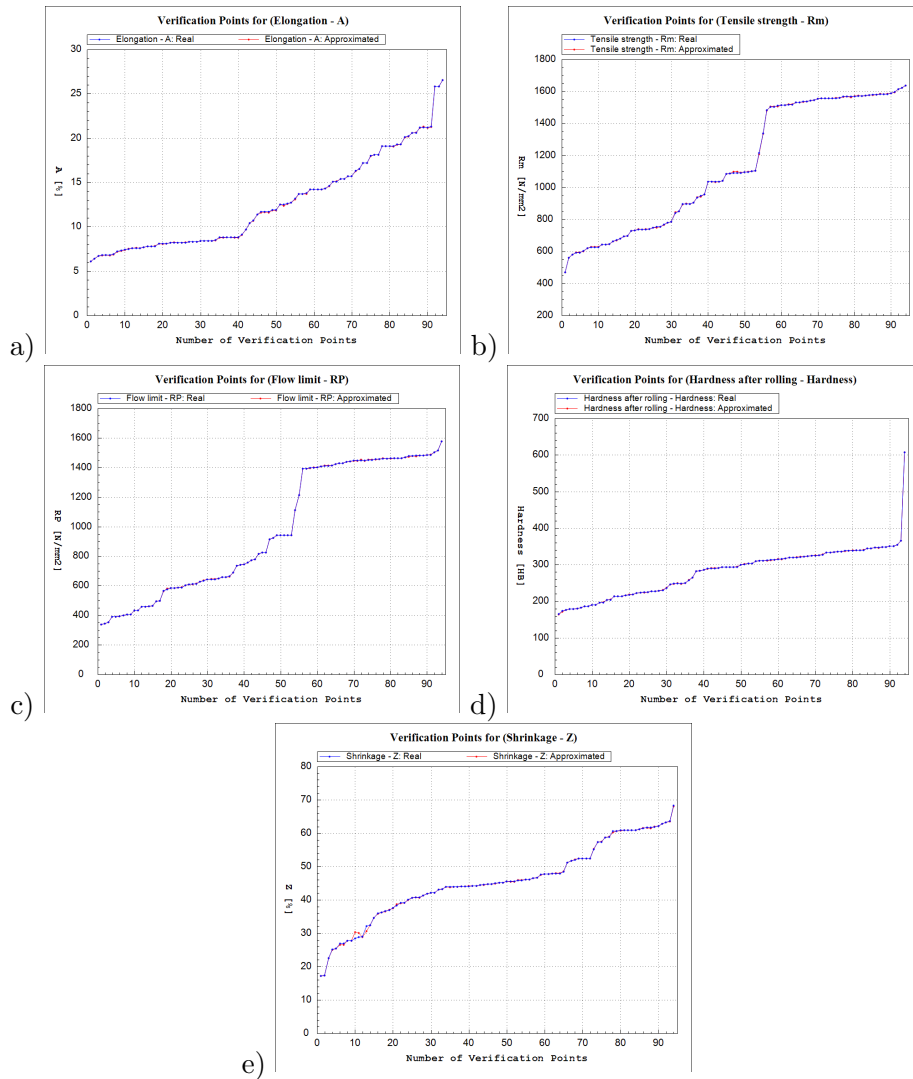
Figure 4: Comparison between actual and approximated values in 94 verification points for elongation (graph a), tensile strength (graph b), flow limit (graph c), hardness after rolling (graph d) and shrinkage (graph e) and sorted from minimum to maximum.

Comparison between actual and approximated response in a number of randomly selected verification points can give us precious indication of the quality of approximation. A problem that we notice is that training data, especially data from real processes are grouped in clusters. This means that chosen verification points are close to some other points from the training
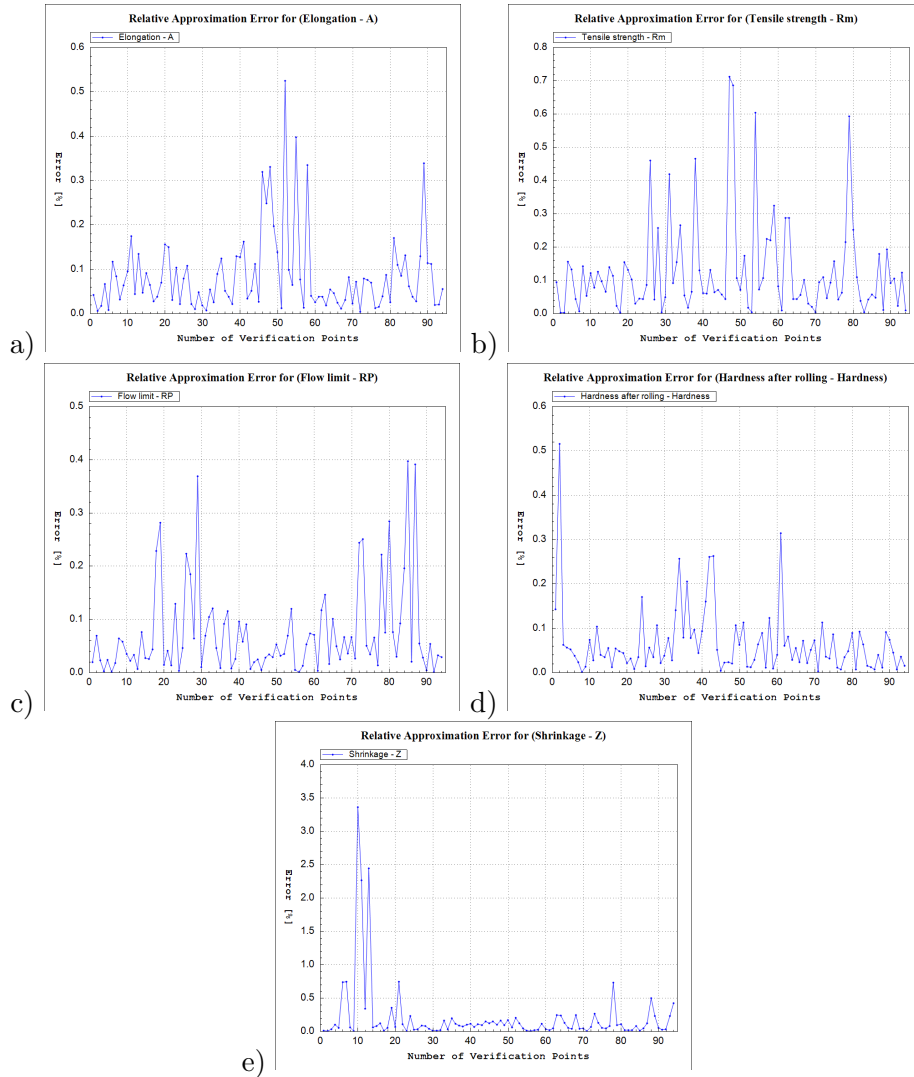
10

Figure 5: Relative errors in 94 verification points for elongation (graph a), tensile strength (graph b), flow limit (graph c), hardness after rolling (graph d) and shrinkage (graph e).

set that remain involved in the training procedure. Therefore the accuracy of approximation in these verification points is better than actual average accuracy, which is affected by regions where training points are scarcely distributed. In this case we do not exactly know what will happen with the approximation between clusters, because we don't have enough information. In the next study we randomly take 4 data sets from the entire data. 2 sets

were chosen among verification points and other 2 from training points. In each chosen set we varied one parameter, for example, concentration of C, while other parameters were fixed. Parameter was varied within the range defined by the minimum and maximum value of that parameter over all dataset used in training. These kind of tests help us find out how the change of one parameter, for example, element C, influences process output quantities of interest such as elongation, tensile strength, flow limit, hardness after rolling and shrinkage. We performed these tests for all 34 input parameters. The influence of variation in concentration of element C on hardness after rolling is shown in Figure 6. From the graph we can see that if we increase carbon (C) concentration, hardness after rolling also increases. This trend is well known in metallurgy.
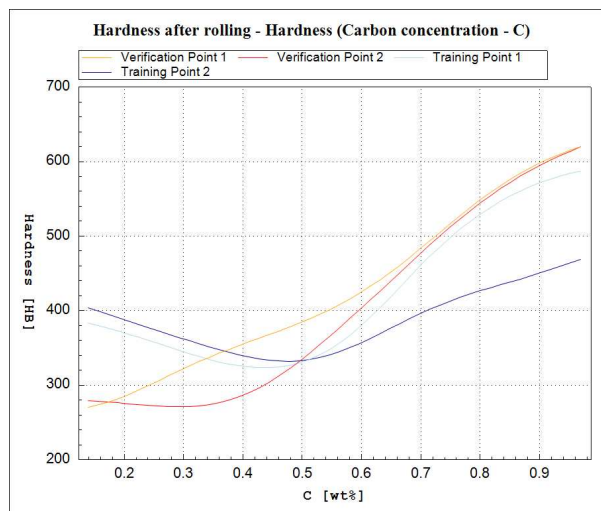

Figure 6: Steel hardness after rolling dependent on carbon mass fraction, calculated by the ANN model for two verifications and two training sets.

In another parametric study we chose two points $(\vec{r_1}, \vec{r_2})$ from the provided data-set, for which the corresponding data sets were included in training of the ANN. Then we took a certain number of equally spaced points on the line segment between these two points (including the chosen points) and effect of variation of a certain parameter were plotted for all points. These intermediate points $(\vec{p_1}, \vec{p_2}, ..., \vec{p_n})$ were calculated according to

$$\vec{p_i} = \vec{r_1} + \left[\frac{\vec{r_2} - \vec{r_1}}{n+1} \right], \tag{2}$$

12

where n is the number of intermediate points. The endpoints were included in the training of the ANN while intermediate points were not. Arrangement of points is schematically shown in Figure 7.
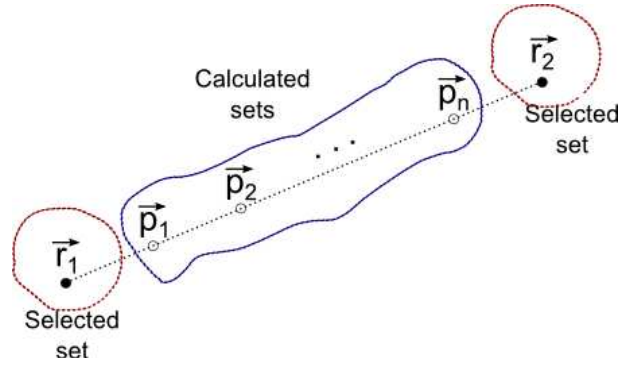

Figure 7: Scheme for parametric study for calculated points on the line $(\vec{p_1}, \vec{p_2}, ..., \vec{p_n})$ between two points chosen from training data $(\vec{r_1}, \vec{r_2})$.

In each of the points from Figure 7 we varied one parameter (concentration of C in this case) over the whole range of values, while other parameters remained unchanged. Influence on hardness is shown in Figure 8.
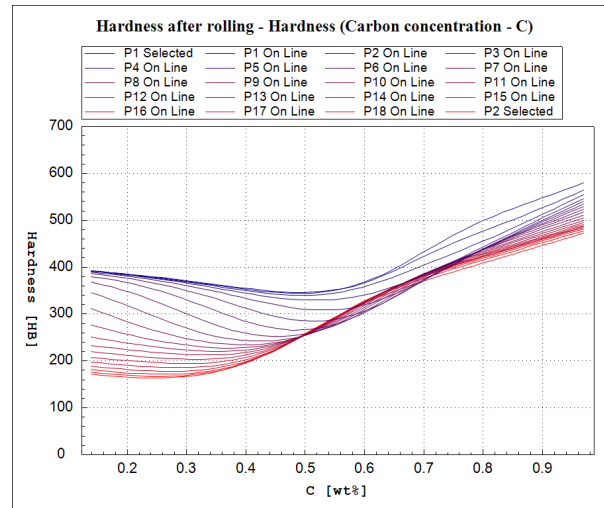

Figure 8: Steel hardness after rolling dependent on Carbon mass fraction, calculated by the ANN model on 2 selected points from the training data and 18 calculated points on the line between them.

13

In this test we can find out how smoothly the curves on the graph pass form point $\vec{r_1}$ to $\vec{r_2}$. Because the points between $\vec{r_1}$ and $\vec{r_2}$ were not included in training, one could expect lower accuracy of the approximated response in these points. In the next study we examine how uniformly the parametric space is covered by the training data. We first calculate for each point its smallest weighted Euclidean distance to any point from the training set.

Figure 9 show distribution of training and verification points according to their distances to the $N+1$-st closest training point where is the number of input parameters. Smaller blue points on the graph represents training points while larger red points represent verification points. We can see from the graph that the distance to the closest points varies a lot, and a large portion of points don't have close neighbours. This means that distribution of data points in parameter space is not very uniform. This could be expected with respect to the fact that data was obtained from the actual industrial line. In steel production, a number of standardized steel qualities are used with narrowly defined chemical compositions, for which process is adjusted according to expert knowledge generated by past experience. Clusters of data points are therefore formed around parameter settings that are commonly in use.
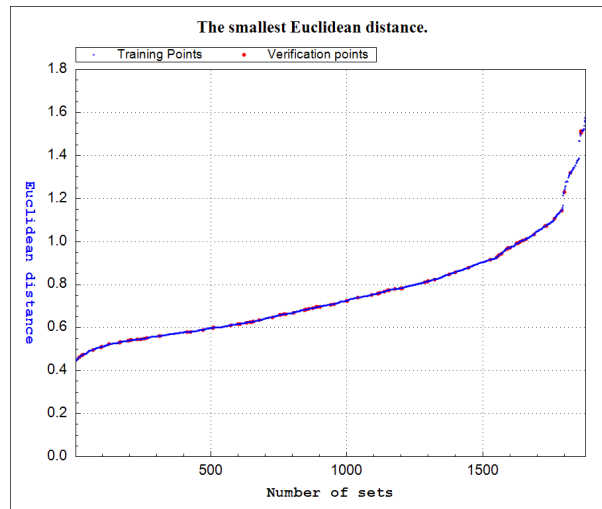

Figure 9: Minimum weighted Euclidean distance from each point of the training data set to the $N+1$-st closest training point and sorted from minimum to maximum.

14

6. Conclusion

A dedicated software framework to support Artificial neural network models was developed. An ANN approximation model for complete steel production process chain was presented. The approximation model was built on basis on 34 process parameters that turn out to be influential. Five output values were modelled which represent important outcomes of the production process. Some parametric studies were performed to examine the accuracy of the approximation. Trends exhibited by approximated response were consistent with metallurgical knowledge and practical experience. However, the accuracy over the whole domain in the parametric space is not yet satisfactory for reliable use in tuning of process parameters. Accuracy varies over domain of interest due to clustering of sampling points contained in the data captured from the industrial production line.

Further development will be directed in development of new methods for assessment of the quality of training data and accuracy of approximation. In particular, meaningful ways of quantitative description of multidimensional distribution of training points in space must be developed and used in of optimal selection verification points. Error estimators will be developed and integrated with optimization and other procedures where the approximate models will be utilized. On the other hand, feedback about critical influential factors is continually sent back to industry in order to improve the accuracy of measurements and constancy of conditions at points that critically affect repeatability of the process.

References

[1] H. K. D. H. Bhadeshia. Neural networks in materials science. *ISIJ Iinternational*, 39:966–979, 1999.

[2] Z. Jančikova, V. Roubiček, and D. Jouchelkova. Application of artificial intelligence methods for prediction of steel mechanical properties. *Metalurgija*, 47(4):339–342, 2008.

[3] P.T. Sampaio, A.P. Braga, and T. Fujii. Neural network thermal model of a ladle furnace. 2009.

[4] T. Martinetz, P. Protzel, O. Gramchow, and G. Sorgel. Neural network control for rolling mills. *EUFIT*, 94:147–152, 1994.

[5] S. Bouhouche, M. Lahreche, and J. Bast. Control of heat transfer in continuous casting process using neural networks. *Acta Automatica Sinica*, 34(6):701–706, 2008.

[6] Store Steel. Product program. 2012. URL `http://www.store-steel.si/ProizvodniProgramE.asp`.

[7] B. Verlinden, J. Driver, and I. Samajdar. Thermo-mechanical processing of metallic materials. 11, 2007.

[8] W.R. Irving. Continuous casting of steel. *The Institute of Materials(UK), 1993,*, page 216, 1993.

[9] J.G. Lenard. Primer on flat rolling. 2007.

[10] R. Vertnik. Heat and fluid flow simulation of the continuous casting of steel by a meshless method. 2010.

[11] R. Vertnik and B. Šarler. Simulation of continuous casting of steel by a meshless technique. *International Journal of Cast Metals Research, 22*, 1(4):311–313, 2009.

[12] I. Grešovnik, T. Kodelja, R. Vertnik, and B. Šarler. A software framework for optimization parameters in material production. *Applied Mechanics and Materialds*, 101-102:838–841, 2012.

[13] Aforge.NET. 2012. URL `http://www.aforgenet.com/`.

[14] NeuronDotNet. 2012. URL `http://neurondotnet.freehostia.com/`.

Measurements of ionospheric properties at low elevation angles using EISCAT incoherent scatter radar

Marko Vučković

University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica

Abstract

EISCAT, the European Incoherent Scatter radar is a powerful incoherent scatter radar which can be used for measuring ionospheric effects like electron density profiles, electron and ion temperature, ion mass, plasma velocity. This paper describes EISCAT based ionospheric measurements along slant ray path, which was performed for the first time. EISCAT radar was set to follow a passing by GPS satellite every five minutes with update of elevation and azimuth angle. We present experimental setup, date processing and the behavior of ionospheric properies.

Keywords: EISCAT radar, electron density profiles, ionospheric properties

Introduction

Classical physics tells us that an individual electron scatters electromagnetic radiation with a cross-section of σ_e=1x10^{-28} m^2. This phenomenon might be used to probe the ionosphere [1]. The first observations of incoherently scattered signals were made using a transmitted power of 6 MW at 41 MHz [2]. This experiment immediately showed the possibility of measuring the electron density at heights well above the F2 peak [3]. It has been demonstrated that measurement of the spectrum of the scattered signal can provide a number of ionospheric parameters, such as: electron concentration N_e, electron temperature T_e, ion temperature T_i, ion drift velocity v_i. All these processes can be measure by EISCAT radar system. A typical EISCAT measurement in the F2 layer involves a scattering volume of 100 km^3, containing a total number of order 10^{23} electrons, the total scattering cross-section is of the order of 10 mm^2. It follows that of the megawatts of power

---

in a transmitted pulse, only micro-watts will be scattered and femtowatts will be received by the antenna. Therefore, powerful transmitters have to be accompanied by large antennas, sensitive receivers and sophisticated signal processing techniques. The firs radars have been constructed to be 'monostatic' and 'multistatic' [4].
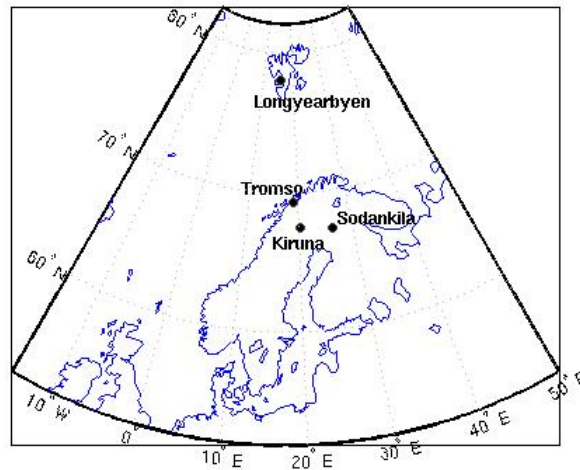

Figure 1: Sketch map of EISCAT. The UHF transmitter is at Tromsø; the receivers are in Tromsø, Kiruna and Sodankilä. The VHF transmitter and receiver are at Tromsø.

EISCAT incorporates both types of radar. Tromsø in the north of Norway is the location for two monostatic radars, one operating at UHF (933,5 MHz) and other at VHF (244 MHz). However, precise timing allows the pulses transmitted from Tromsø at UHF to be received at two remote sites, Kiruna in Sweden and Sodankilä in Finland, which enables three independent components of ion velocity to be measured at any point in the ionosphere accessible to all three antennas [5]. The simplest explanation of incoherent scatter is that of an electromagnetic wave randomly scattered by the individual electrons of the ionospheric plasma. This explanation assumes that the electrons are totally independent so that the strength of the scattered signal is proportional to the electron density [5]. However, this simple explanation also suggests that the scattered radiation will have a Doppler-broadened spectrum resulting from the random thermal velocities of the electron population, with a mean Doppler-shift corresponding to the systematic or 'bulk' motion of the electrons [2,6].

Experimental setup

The data used in this analysis were collected with the EISCAT radars placed in Tromsø (69°35′ N, 19°14′ E), on Svalbard (78°09′ N, 16°1′ E) and Sodankilä (67°2′ N, 26°38′ E), see Figure 1. The EISCAT radars use a frequency of 931 MHz and 500 MHz and maximum power of 1.3 MW. The Tromsø antenna is a steerable 32 m UHF parabolic antenna with azimuth angle range from $-270°$ to $270°$ and elevation angle range from $30°$ to $90°$. For the analyses presented in this paper, we used unprocessed (raw) data obtaned from the EISCAT system and stored in the form of binary files.

Table 1: List of the multiplying factors (Magic constant) used to calibrate the data and to monitor day-to-day changes for for both bella and folke experiments between 13 and 16 December 2011.

Day	Experiment	Magic const.
12-Dec-11	bella	1.07
	folke	/[1]
13-Dec-11	bella	0.83
	folke	0.89
14-Dec-11	bella	0.79
	folke	2.29
15-Dec-11	bella	0.79
	folke	0.89
16-Dec-11	bella	0.84
	folke	0.46

There are a few types of experimental activities using the EISCAT radars. We focused on the experiment dedicated to the measurement of E and F region height (bella experiment). In our experiment, the Tromsø and Svalbard antennas are fixed and parallel. The remote sites make continuous measurements at the E and F region height within range of 70 to 700 km. The second experiment that we focused on (folke experiment) is dedicated to meridional scanning. The temporal resolution of the experiment was in our case about 6 sec. From these measurements, we obtained electron dencity profiles up to the height of 700 km. To decrease measurement errors

---

[1]We didn't have the coordinates of GPS satellite for Svalbard.

electron dencity profiles are obtained as an average value of the integrated along the event. We investigated the cases with 3 different integration constants, 300 sec (entire event), 60 sec and 150 sec. The values were used for fitting electron densities and compering with ionosonde readings can be seen in Table 1. Default fit parameters and heights for each incoherent scatter radar can be seen in Table 2. The parameters used in our experiment are: critical frequency foF2[2], altitude range (default 90-140, 180-500), maximum error allowed (default 1) and minimum elevation allowed (default 75) respectively. The details and figures will be discussed in the next section.

Table 2: Default fit parameters and heights used for the processing of bella and folke experimental data between 13 and 16 December 2011.

Parameter	UHF (930 MHz)	VHF (224 MHz)	ESR (500 MHz)
Electron density	All heights	All heights	All heights
Ion temperature	Above 80 km	Above 100 km	Above 90 km
Temperature ratio	107-1500 km	120-1500 km	113-1500 km
Collision frequency	90-107 km	Never	Never
Ion drift velocity	All heights	All heights	All heights
Ion composition	Never	Never	Never

Results

During our EISCAT measurement campaign, which last 5 days, from 12-16 of December 2011, we noticed that at the first two days (12th and 13th Dec. 2011) we had very active conditions and at the last three days we had almost quite conditions. Because of that, we decided to present electron profile dencities taken from day with very activ conditions (13th) and another taken from day with almost quite conditions (16th). What can be seen from Figures 2-4 show the standard colour plots of electron density profiles of the analysed EISCAT UHF data for 13th and 16th December 2011.

---

[2]foF2 is the critical frequency (the limiting radio frequency) below which a radio wave is reflected by, and above which it penetrates and passes through the ionized medium at vertical incidence.
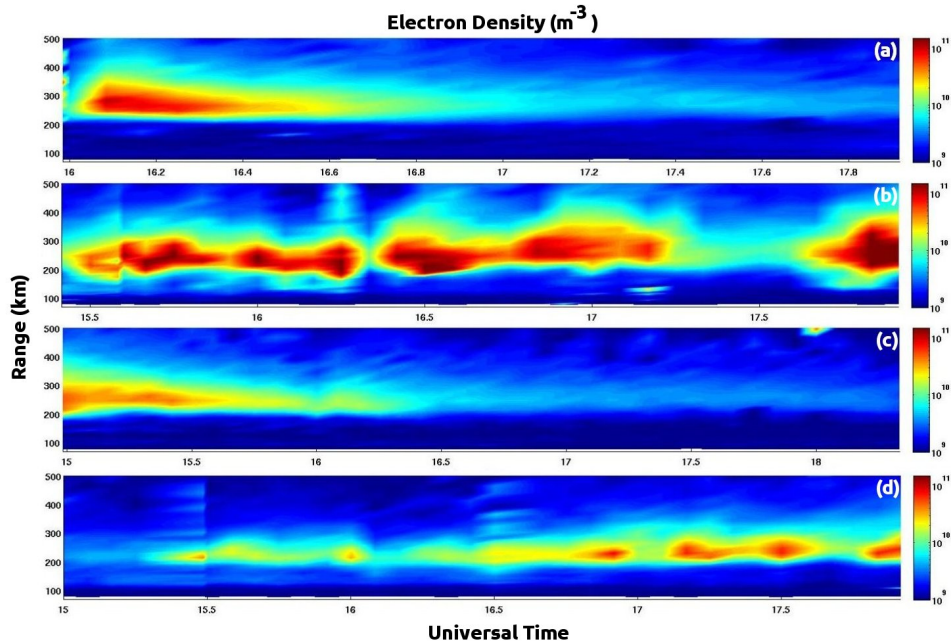
Figure 2: Electron density obtained using a default multiplying factor used to derive the densities from the raw data. (a) Electron density profile at Tromsø on 13th Dec., (b) Electron density profile at Svalbard on 13th Dec. (very active conditions). (c) Electron density profile at Tromsø on 16th Dec., (d) Electron density profile at Svalbard on 16th Dec. (almost quite conditions but the region is still active). Temporal results of electron densiti profiles is 5 minutes.

It can be noticed two or more peaks, which means that the F region is active after maximum peak. The more detailes can be seen in Figure 3 and 4 according to integration time of 60 and 150 sec over 5 minutes interval. However, we won't discuss here about perturbations and effects of this phenomenon. What is interesting to mention is that Figure 2 is obtained using a default multiplying factor used to derive the densities from the raw data. From Figure 3 and 4 it can be seen how the electron density profiles change according to integration time. What is more, the new multiplying factors (Magic_constant) have been used to calibrate the data and to monitor possible changes.
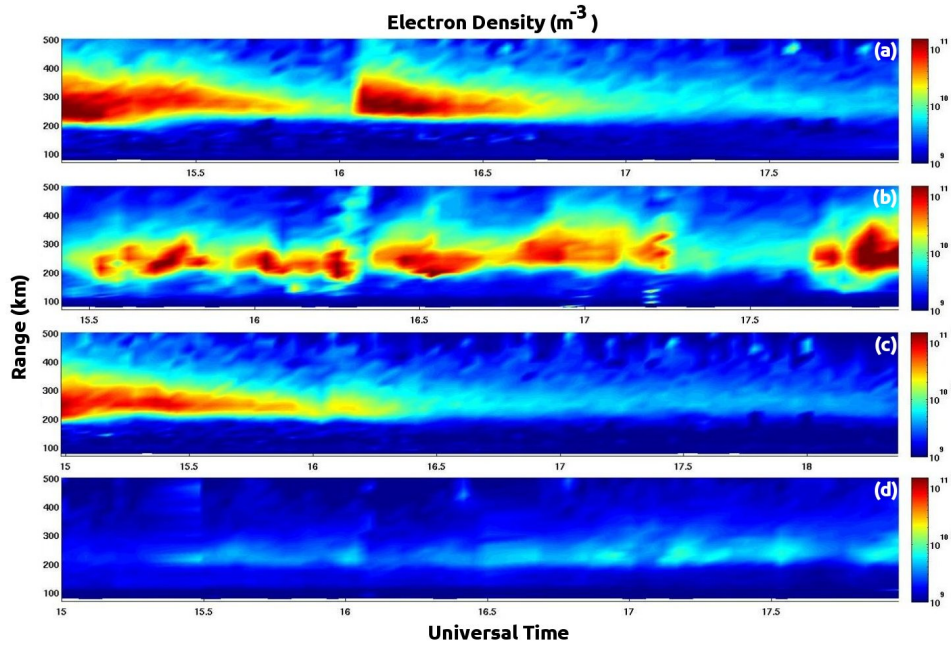
5

Figure 3: Electron density observed with integration time of 150 sec and with a calibration factor from Table 1. (a) Electron density profile at Tromsø on 13th Dec., (b) Electron density profile at Svalbard on 13th Dec. (c) Electron density profile at Tromsø on 16th Dec., (d) Electron density profile at Svalbard on 16th Dec. All data are averaged over integration interval of 150 sec.

While data recorded at the basic time resolution of the EISCAT systems can be interesting for the study of high time-resolution phenomena, they are also often too noisy for purposes such as data analysis and therefore have to be post-integrated up to time scales of 1 minute or greater. Because of this, we use diferent integration intervals to split up the data for further analysis. After all mentioned calibrations it can be seen (Figure 3 and 4) that the influence of multiplying factor can be neglected but changing the integration time can be used for observing the electron density profiles in a different volumes.
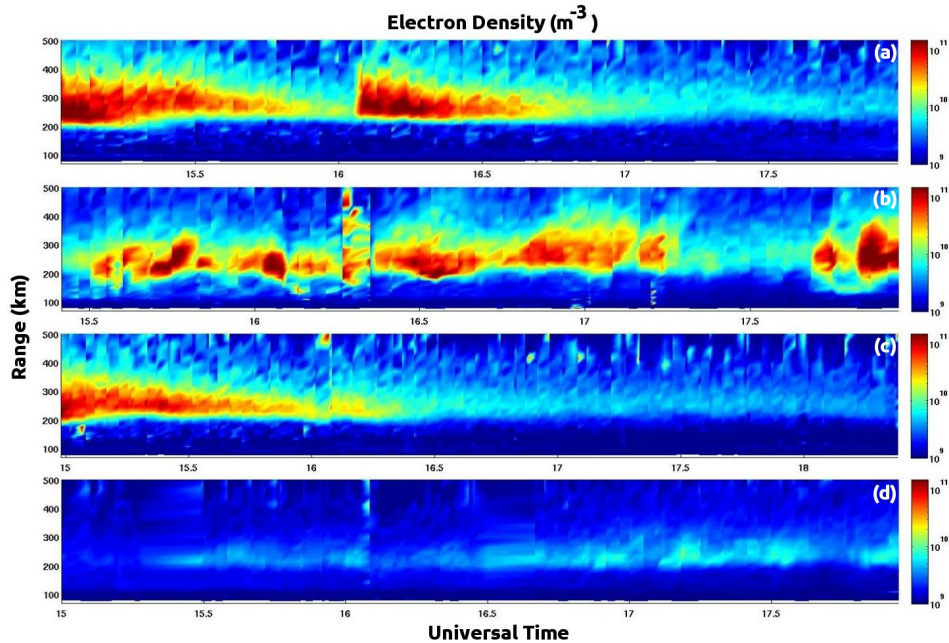
Figure 4: Electron density observed with integration time of 60 sec and with a calibration factor from Table 1. (a) Electron density profile at Tromsø on 13th Dec., (b) Electron density profile at Svalbard on 13th Dec. (c) Electron density profile at Tromsø on 16th Dec., (d) Electron density profile at Svalbard on 16th Dec. All data are averaged over integration interval of 60 sec.

Discussion and conclusions

Presented calibration technique can be used to compare data from dynasonde and the EISCAT radar if they observe the same volume of the ionosphere, which means they have a cross section. It has been shown that calibration methods can not be used for data obtained at very slant ray paths, since the distance between ionospheric volumes is too large. In other words, the geometry of ionosonde and radar are not overlap. This can be seen if we compare Figure 2 with other figures. In addition to the dynasonde, we plan to compare EISCAT radars to the localed GPS receivers data. Future studies of ionospheric disturbances, sporadic E and F layers as well as polar irregularities caused by magnetic storms will be investigated in future measurement campaigns EISCAT Scientific Association.

References

[1] C. Fabry, *Comptes Rendus de l Academie des Sciences*, Paris, 187, 777, (1982).

[2] K. Bowles, *Physics Review Letters*, I, 454, 1958.

[3] R. W. Schunk & A. F. Nagy, *Physics, Plasma Physics and Chemistry*, Second Edition, 31, (2009).

[4] W.J.B. Baynon & P.J.S. Williams, *Report Progress Physics*, 41, 909, (1978).

[5] H. Rishbeth & P.J.S. Williams, *Q.J. Royal Astronomical Society*, 26, 478-512, (1958).

[6] K. Bowles, *J. Research National Bureau of Standards*, 65D, 1, (1961).

[7] H. Rishbeth & P.J.S. Williams, *Q.J. Royal Astronomical Society*, 26, 480, (1958).

[8] Ingemar Häggström, *Running EISCAT mainland*, Octobar (2008).

Comparison of Total Electron Content measurements by EISCAT radar and GPS ionospheric monitors at high latitudes

Đorđe Stevanović

University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica

Abstract

Irregularities, occuring mainly at equatorial and high-latitude regions, can cause formation of various phenomena in the ionosphere, which can lead to fluctuations in amplitude and phase of GPS signals. This fluctuations are known as scintillations and they are described with various parameters. For the purpose of measuring and calculating this parameters we have used data got from EISCAT Incoherent Scatter Radar (ISR) and GPS Total Electron Content (TEC) and scintillations monitors at Longyearbyen, Svalbard and Tromsø. Comparison of results will lead to improved understanding of small scale irregularities, modification of existing ionospheric models and/or introduction of a new model for prediction of electron plasma irregularities.

Keywords: incoherent scatter radar, GPS monitor, total electron content, ionospheric irregularities

1. Introduction

Solar plasma continuously flows from the Sun's surface as solar wind and on its way towards the Earth interacts with various interplanetary magnetic fields, which change its speed and pressure. When solar plasma reaches Earth, it interacts with its plasma and magnetic field, part of its atmosphere called magnetosphere, shaping it in form of compression on the day side and elongated tail on the night side of the Earth. Result of this interaction between Earth's and Sun's plasma are different layers in the

---

atmosphere, one of them is the ionosphere, which effects and characteristics are investigated in this paper. The ionosphere is ionized portion of the upper atmosphere (extending from approximately 60 to around 1000 km), where free electrons occur in sufficient density that they have an significant influence on the propagation of electromagnetic waves [1]. Density of free electrons and ions influence variation in the refractive index when signal penetrates ionosphere. Number of electrons in the ionosphere vary greatly with time, season, solar cycle, geographical location, magnetic activity and with certain solar-related ionospheric disturbances (see Fig. 1).
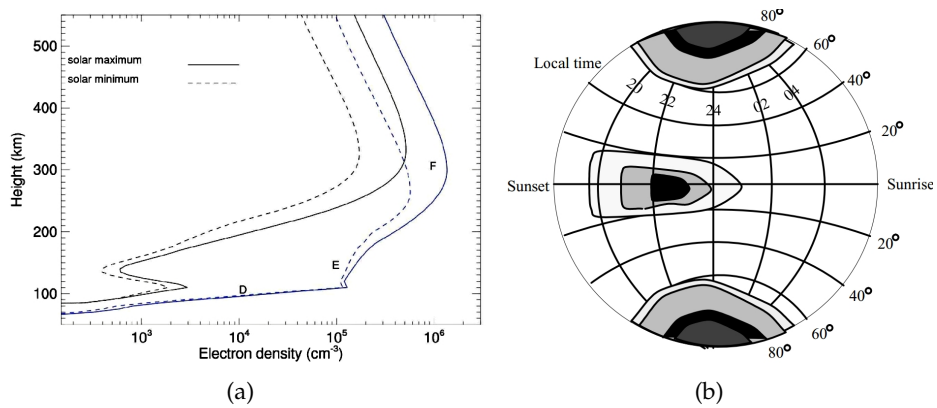

Figure 1: (a) Ion density layers in the ionosphere and their appearance during day(higher values) and night(lower values) and during solar maximum and minimum [2] and (b) schematic presentation of global depth of scintillation fading [3].

Unmodulated variability in the ionosphere has placed significant limitations on the accuracy of navigation applications. Signal delay due to refraction through the ionosphere is unpredictable and most variable source of positioning error for commercial single and sometimes dual frequency receivers, which are much more resistant on this kind of effects, but not immune [4]. This is the main reason why radio scientists are trying for decades to track, record and forecast the electron density distribution in the Earth's ionosphere. Major ionospheric effects on GPS signal are two types of scintillations: refraction or range errors (group delay and carrier phase advance) and diffraction (the radio wave reaches the receiver through multiple paths). These effects are more evident at low and high latitudes. Both types of effects originate in the group delay and phase advance of the GPS signal when it interacts with free electrons along its transmission path. The number of free electrons present along path between two points, in our case between satellite and receiver, in a column of one square meter cross-

section is called Total Electron Content (TEC) [5]. TEC is usually low at night and maximum at round early afternoon, when the Sun's position is approximately few hours past zenith. The extent of this time shift depends on the time that light needs to ionize the layer to the maximum. The high latitude ionosphere can be considerably altered in composition and structure by the various physical processes, such as, ionization and heating processes that effects particle precipitation to be superimposed on photon fluxes. Measurements of TEC values has become available from both space and ground based observations over the last decade. One of the most used method for investigation of ionospheric irregularities is relatively accessible ground based GPS Ionospheric Scintillation and TEC Monitor (GISTM) for collecting and processing received data from GPS signals. Advantage of satellites is that large segments of the ionosphere can be investigated in a short time interval, due to the satellite motion. Second method, one of the most advanced, uses incoherent scattering of signals to measure a set of ionospheric parameters. Main principle of incoherent scattering relays on Thomson scattering, which shows that free electrons are capable of scattering electromagnetic radiation in a particular direction with specified scattering cross section $\sigma = 4\pi \left(r_e \sin\gamma\right)^2$ [6]. Modern theories showed that that ionosphere is composed of electrons which could not be considered like free, due to the fact that they are dependent of electrostatic forces between electrons and positive ions. Spectral width is controlled by the thermal velocities of the bigger and slower ions, until radar wavelength λ is greater than the Debeye length λ_D, which is the scale over which dynamic charge particles screen out electric fields in the ionosphere:

$$\lambda_D = \left(\frac{\varepsilon_0 k T_e}{N_e e^2}\right)^{1/2}, \tag{1}$$

where ε_0 represents permittivity, k is Boltzmann's constant, N_e electron density and e is charge of an electron.

Relationship between this two wavelengths could be represented by constant $\alpha = 4\pi\frac{\lambda_D}{\lambda}$, where if $\alpha < 1$ then scattered signal is influenced by ion-acoustic waves and ion-line takes a look of double-humped spectrum characteristic like in Fig. 2. From information obtained by observing characteristics of spectrum it is possible to calculate ionospheric parameters - total scattered power, total width of spectrum, centre frequency and ratio of the peaks to the trough in the spectrum. Spectrum ion line characteristic characteristic vary with change of electron and ion temperature, content of different ions and ion velocity. More detailed description of all parameters, their derivatives and influence of various factors on signal and system
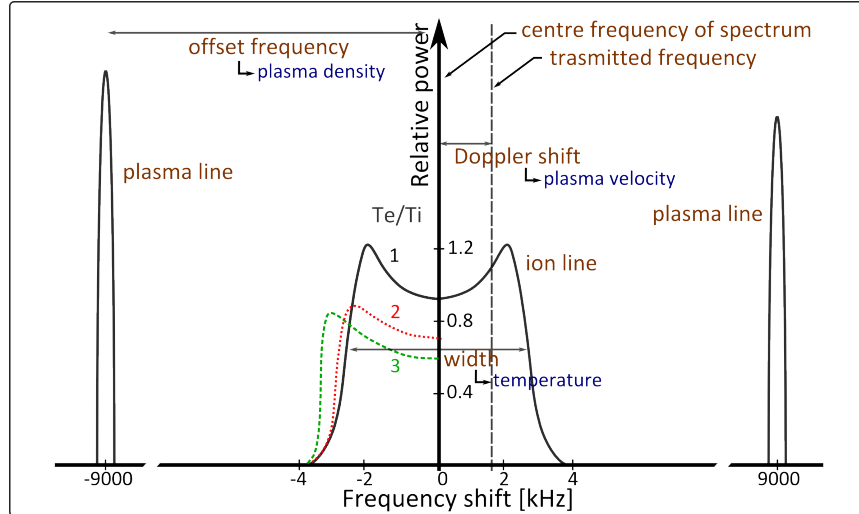
Figure 2: Representation of the incoherent backscatter power spectrum with ion line and plasma lines and derived parameters. Dotted lines represents changes of ion-line characteristics with dependance on relationship between electron and ion temperature.

could be found in [6, 7, 8]. In this paper the focus will be on ground based measurements with two different technique, ISR and GISTM, used during measurement campaign, which will be described in following sections.

2. Experimental setup

During measurement campaign, which was carried out between 13th and 16th December 2011, we have used one of the most advanced instruments for measuring dynamic effects in the ionosphere, so called incoherent scatter radars (ISR) placed in Tromsø Norway and Longyearbyen, Svalbard. The TEC values, usually used for communication applications, are deduced from GPS measurements and represent a sum of all the electrons up to 22 000 km, which is much higher than can be reached with the radar. The TEC integrated up to 700 km represents roughly half of the total integrated TEC. The largest effects of particle precipitation and electric fields are in the E and F regions of the ionosphere, which are covered by the radar. This fact was the lead of project to check is it possible to make a study of the relative behaviour of the TEC with the radar data processed in the appropriate way and compared with values from GPS monitors. Two EISCAT transmitters and GISTM GSV4004B phase and amplitude scintillation receivers were deployed in Northern Europe during measurements

4

from 13th to 16th December 2011. Measurements of ionospheric parameters like phase and amplitude scintillation and TEC from GISTM receivers and electron density, ion and electron velocity and temperature have been archived and analysed in purpose to compare TEC values from both devices. Measurements with ISR have been done in following way, GPS satellites have been tracked every five minutes by moving antenna dish in accordance with known coordinates of satellites movements. When elevation angle become very low, antenna dish was pointed to the next satellite, in our case first satellite was PRN 20 and second one was PRN 23. Radar scanning mode which was used during measurements for tracking GPS satellites was modification of CP1 scanning pattern mode (an experiment using many different pulses and pulsecodes to probe along the fieldline above ISR, IP1 mode. In this mode, the Tromsø and Svalbard antennas are fixed parallel to the magnetic field, so all data is altitude versus time along the field-line. The remote sites made continuous measurements at the F-region height with radar range from 70 to 700 km. From those measurements, we computed the electron content up to 500 km height and the F region parameters. Within this project we get the data of GISTM monitors set on Svalbard and in Tromsø for observing variations in the satellite signals caused by irregularities of various dimensions.

2.1. EISCAT radars

The European Incoherent SCATter (EISCAT) radar system represents advanced geophysical tool for investigation the atmosphere by using various scanning methods. The system is composed from very powerful transmitters, which uses signals of order of megawatts, and very precise and sensitive low-noise receivers, which collect coded backscatter signals of order of picowatts. EISCAT receivers are located in Kiruna (Sweden), Sodankylä (Finland), Tromsø (Norway) and Longyearbyen, Svalbard, and transmitters are placed in Tromsø and Longyearbyen, Svalbard (see Fig. 3). There are two basic types of formation of receivers and transmitters:

1. monostatic - transmitter and receiver are on the same site (Tromsø and Svalbard),
2. multistatic - transmitter and receiver are separated (Tromsø together with receivers in Kiruna and Sodankylä).

With multistatic type it is possible to get more information about ionosphere at different ground places (receivers) from any point in the ionosphere accessible to the antennas. Investigation of scattering mechanism during last century showed that it is possible to get a number of ionospheric

Figure 3: Geographic positions of the EISCAT radars (modification of figure in [9]). Receivers used for tristatic measurements are represented in green color, while transmitter/receiver sites are red.

parameters from measurement of received spectrum of the backscattered signal. List of parameters, which are possible to get directly or indirectly, are presented in Table 1. From this parameters it is possible to derive valuable information about electric field, neutral wind and temperature, photoelectron energy and the composition of the atmosphere which is very difficult to investigate at ionospheric heights.

2.2. GPS monitors

Another measuring technique, popular in analysing the ionosphere by monitoring and tracking GPS signals from ground, consists of special receiver made for monitoring and recording ionospheric processes. Data recorded by one of this receivers, GPS Ionospheric Scintillation and TEC Monitor (GISTM) system (model GSV4004B), was used in this paper for

6

Table 1: List of ionospheric parameters accessible by incoherent scatter radar measurements

Measured	Derived indirectly
Electron density, N_e	Neutral particle density, ϱ
Electron temperature, T_e	Neutral gas temperature, T
Ion temperature, T_i	Neutral air velocity, U
Ion composition	Pedersen and Hall conductivity,
Ion drift velocity, v_i	Σ_P, Σ_H
Ion-neutral collision frequency, v_{in}	Electric current
Photoelectron flux	Heat flux

comparison with ISR analysed data. The GISTM is capable of measuring TEC and variations in amplitude and phase scintillations of GPS signals in real time. The GISTM is dual-frequency GPS receiver with an L1/L2 GPS antenna and 12 channels for tracking up to 11 GPS (and GLONASS) satellites and 2 Satellite-Based Augmentation System (SBAS) satellites, and a low phase noise oven-controlled crystal oscillator (OCXO) for monitoring phase scintillation [10]. The GISTM has wide-bandwidth tracking loops to maintain lock longer during intervals of strong ionospheric scintillation, and samples at a rate of $50\,Hz$ and code/carrier divergence at $1\,s$ ($1\,Hz$) to calculate the scintillation statistics S_4, TEC and $\sigma_{\phi 60}$ [11]. The primary purpose of the GISTM system is to collect ionospheric scintillation information and store these statistical data on the receiver controller hard disk for further analysis. These data were used to reconstruct the scintillation parameters of part of the ionosphere above locations near to EISCAT ISR's and compare data from both sources. TEC values was calculated from combined L_1($1575.42\,MHz$) and L_2($1227.60\,MHz$) pseudo-range and carrier phase measurement.

3. Results

With the purpose to compare measurements with ISR and recorded data from GISTM, we will use the most interesting results from 14th December, when we recorded disturbances in the ionosphere (see upper right graph in Fig. 4), and data recorded on 16th December like reference, this was the most quietest day. Fig. 4 represents analysed raw data for electron density for radar range of $700\,km$ pointing in direction of tracked satellite. Small jump is noticeable around 16 hours UT, which appears due to change of antenna direction from first tracked satellite, PRN 20, to second one, PRN 23. Disturbances in electron density characteristic of F layer, which appears

in the upper right graph in Fig. 4, are probably caused by particle precipitation, polar cap phenomena when energetic particles can enter deep in to the Earth's atmosphere and react with charged particles in the ionosphere. Measurements were done during very quiet geomagenetic period, during which global k_p and D_{st} indexes were around zero, through whole measurement campaign. During processing data, we have meet with series
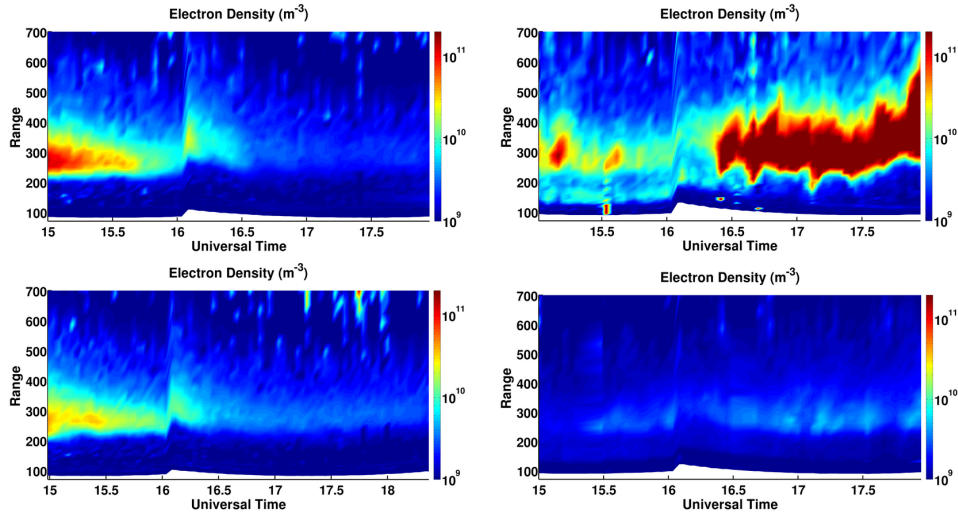

Figure 4: Measurements of electron density from Tromsø and Svalbard ISR. First raw of figures represents data from Tromsø and Svalbard, respectively, measured on 14th December. Second raw are results from quiet ionosphere at 16th December.

of calibrations of raw data, which is needed for getting final comparable values. TEC values for ISR are calculated by integration electron content values with trapezium rule, which appears like most suitable solution. Integration boundaries took place from 70 km to 500 km, due to the fact that comparable values above 500 km and are influenced by internal instrument errors. This values were shifted in time, because radar integration time and integration time of GPS monitors was different. Radar's integrated values was taken from an interval time equal to 2.5 minutes, which included the time that the radar need to point the antenna, so the real integration time was less than 2.5 minutes. The reference time for each value lies in this interval, so we choose middle value as our reference. In the end, we had the GISTM TEC values with their own times and the ISR TEC values where each TEC is referenced to a time which is in the middle of each measurements interval time (around 2.5 minutes). Once we had the radar TEC time

8

axis, we take the GISTM TECs inside the same ISR interval (2.5 min) and made an average of them, with reference time of the ISR. Normalization of TEC values for GISTM, so that it become aligned and comparable with radar TEC, are presented on Fig. 5 and Fig. 6. Figures represents comparison of final analyzed and corrected data, together with GISTM S_4 and $\sigma_{\phi 60}$ data.
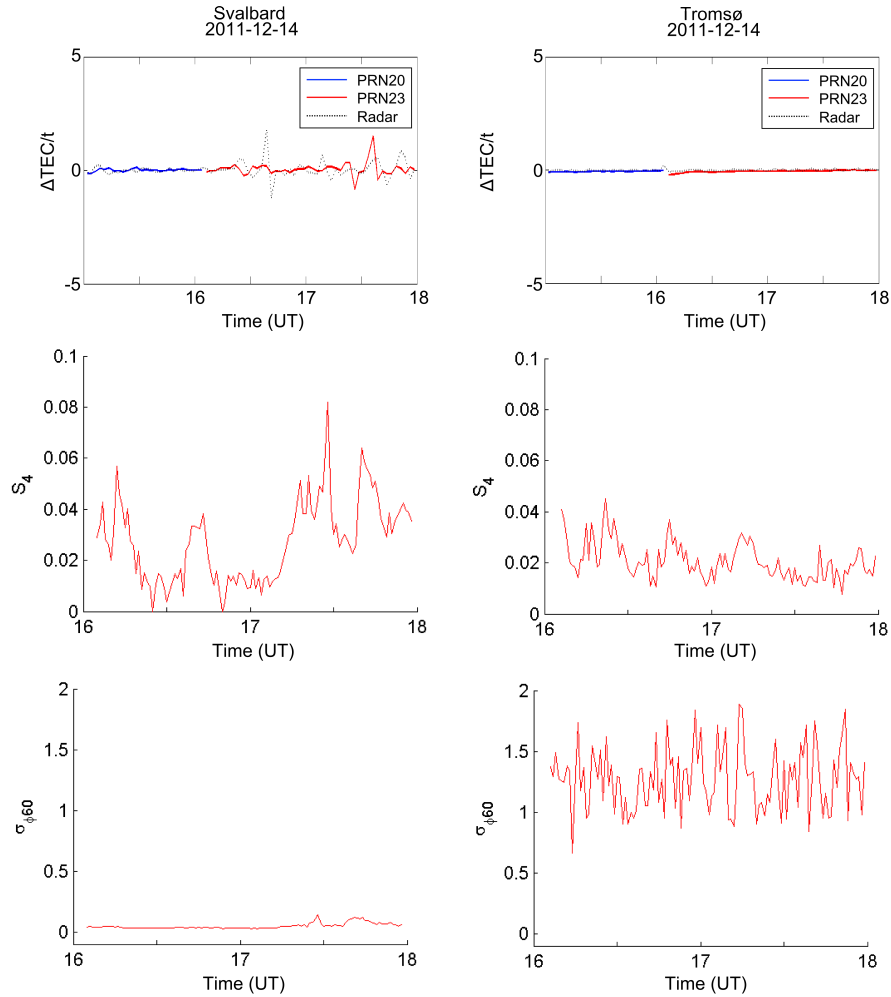

Figure 5: Comparison of TEC values from ISR and GISTM and S_4 and $\sigma_{\phi 60}$ from GISTM for Svalbard and Tromsø, respectively, measured on 14th December. Disturbances occur in last of characteristic for TEC and S_4, probably due to particle precipitation. Time refers to universal time.

Comparing results for TEC values it could be seen that characteristics

mainly coincide with each other. Deviations are noticeable at last part of TEC characteristic for 14th December, probably due to particle precipitation influence caused variations in electron density. S_4 and $\sigma_{\phi 60}$ characteristics are given to show that fluctuations existed in amplitude while phase had normal values, but it could be seen that deviations are very high in $\sigma_{\phi 60}$ characteristics for Tromsø for 14th and 16th December due to instrument error.
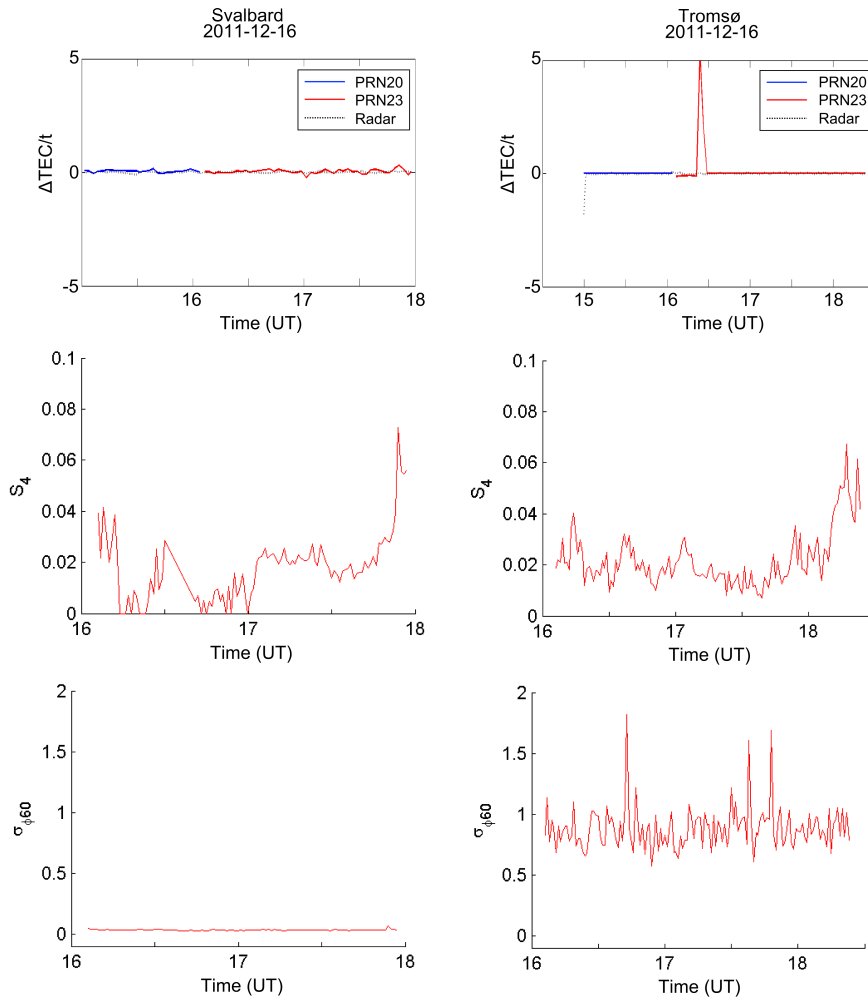

Figure 6: Comparison of TEC values from ISR and GISTM and S_4 and $\sigma_{\phi 60}$ from GISTM for Svalbard and Tromsø, respectively, measured on 16th December. Very quiet period, disturbances occur in TEC measurements with GISTM probably due to instrument errors. Time refers to universal time.

10

4. Conclusions

In this paper it has been shown that results of coordinated GISTM and ISR measurements of ionospheric properties are in good agreement. In order to obtain comparable values of both systems, calibration of ISR data was performed together with normalization and alignment of GISTM data. Comparison of results from two different locations, Svalbard and Tromsø, showed that disturbances on 14th December were localised, with significant influence on the ionosphere above measuring sites on Svalbard. Instead of TEC depletion with time, TEC had increased and reached its maximum value at about 18:00 UT. Disturbances in the ionosphere during this period caused small deviations in TEC values calculated by ISR and GISTM, which exist due to the fact that different methods have been used for measurements. Broader spectra of measuring instruments set on Svalbard could be used for further investigation on ionospheric phenomena during 14th December, but also as a reference for precision and review of all instruments used for this and similar kind of investigations.

The presented methodology will be used as a basis for further investigation of existing TEC models and for the devising new models for small scale irregularities. Long baseline EISCAT measurements can be used for investigation of slow processes in high latitude ionosphere. Investigation of polar irregularities like polar arcs and patches, but also traveling ionospheric disturbances and sporadic E layers, could be analysed in details. Further investigations would be directed on distribution of physical model of ionospheric irregularities, which goal would be to support mitigation techniques of bad influence of ionospheric irregularities on global navigation satellite system signals and point out weak points of commercial receivers, in order to improve their accuracy.

References

[1] R. W. Shunk and A. F. Nagy, *Ionospheres: Physics, Plasma Physics, and Chemistry*, First Edition, 2000.

[2] E. E. P. Josan, *Latitudinal (apex-height) variations of ion drifts in the ionosphere at low- and mid-latitudes*, Thesis (Ph.D.)- The University of Texas at Dallas, 2009.

[3] D. Anderson and T. Fuller-Rowell, *The Ionosphere*, Space Environment Center, 1999.

[4] G. X. Gao, S. Datta-Barua, T. Walter and P. Enge, *Ionosphere Effects for Wideband GNSS Signals*, ION Annual Meeting 2007, Cambridge, Massachusetts, April 2007.

[5] `http://www.insidegnss.com/node/1579`, accessed on 2nd February 2012.

[6] W. J. G. Beynon and P. J. S. Williams, *Incoherent scatter of radio waves from the ionosphere*, Rep. Prog. Phys. 41 909, 1978.

[7] H. Rishbeth and P. J. S. Williams, *The EISCAT Ionospheric Radar: the System and its Early Results*, Quart. J. Royal Astr. Soc., 26, 478-512, 1985.

[8] D. Alcaydé, *Incohent Scatter Theory, Practice and Science*, Collection of lectures given in Cargese, Corsica, 1997, EISCAT Tech Rep 97/53, 1995.

[9] `http://mapsof.net/map/norwegian-sea-blank-map`, accessed on 4th May 2012.

[10] A. J. Van Dierendock, J. Klobuchar and Q. Hua, *Ionospheric Scintillation Monitoring Using Commercial Single Frequency C/A Code Reciever*, ION GPS-93 proceedings, Salt Lake City, pp.1333 âĂŞ 1342, 1993.

[11] GSV4004B, *GPS Ionospheric Scintillation & TEC Monitor, (GISTM)*, User's manual, August 2007.

12

Photodegradation of chlorantraniliprole and characterization of its transformation products

Vesna Lavtižar

University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica

Abstract

A photostability of a new and already widely used insecticide, chlorantraniliprole, was studied in water solution by simulated solar irradiation. For this paper, the chlorantraniliprole's photodegradation pathway has been investigated for the first time, with characterization of the four main degradation products. Two derivatives of chlorantraniliprole are formed exclusively photochemically, while one is formed in water or basic media by spontaneous transformation of the first photodegradation product. The fourth transformation product is formed by hydrolysis of chlorantraniliprole in basic water. The degradation pathway is very dependent on the acidity of the media.

Keywords: chlorantraniliprole, photodegradation, transformation products, photolysis

1. Introduction

In the environment, pesticides can be degraded in several ways, forming a variety of different transformation products. There is an increasing concern regarding their presence in the environment. Early studies demonstrated that certain transformation products of pesticides can be more toxic than the parent compounds [1], more persistent [2] and can in addition possess properties that enable them to enter environmental areas not accessible to the original substances (ie. ground water) [3]. One of the substances whose the environmental impact of its decay products needs to be studied in detail, is chlorantraniliprole [1] (CAP), a newly developed insecticide, belonging to the anthranillic diamide chemical group [4]. For its high insecticidal efficiency, mammalian safety and its low toxicity to most of the non-target species, CAP is becoming widely used for pest control in orchards, vineyards and on crops and is eligible also for use in integrated pest

---

[1] trade name Rynaxypyr®

management programs [4]. Most of the research is focusing on the insecticidal efficiency and modes of action on pests [4]-[10], independent studies of the behavior of CAP in the environment are rare. These independent studies mostly describe the use of analytical methods for the determination of CAP [13, 14]. Review of scientific literature indicates a clear lack of results regarding the stability of CAP and its degradation products, their mobility in the environment, toxicity to non-target organisms and risk to humans.

To some extent, the photostability of CAP was investigated by the producers (data published in Food and agriculture's report) [12] and Enviromental protection agency [11], which claim it to be extremely stable in the environment. Major paths of dissipation are expected to be alkaline-catalyzed hydrolysis, photodegradation in water, leaching and runoff. They reported that photolysis of radiolabelled ^{14}C-chlorantraniliprole in sterile aqueous buffer solution (pH 7.0) yielded three main degradation products for which the chemical structure was given. However, in the above mentioned technical reports, the degradation pathways are poorly established, transformation intermediates are inadequately described and the descriptions of experimental procedures are more or less unavailable. In this paper, the photdegradation of CAP in aqueous media including the experimental setup and the characterization of its four transformation products is described in detail.

2. Experimental setup

2.1. Analytical and experimental procedures

Solvents and reagents (Aldrich, Fluka, Merck) were of analytical and HPLC grade and were used as purchased. The presence of chlorantraniliprole[2] and its degradation products was being followed using high pressure liquid chromatography (HPLC) based on the diode-array detector[3] and Luna C18 column[4]. Mobile phase consisted of acetonitrile (MeCN) and acidified ddH$_2$O (0.1 % formic acid) in the ratio 60 : 40, flow rate was 1 mL/min, injection sample volume was 30 μL and the column was kept at constant temperature of 22 °C. Compounds were characterized by ^{1}H and ^{13}C nuclear magnetic resonance (NMR)[5], elemental analysis and liquid chromatography coupled with time-of-flight mass spectrometry. Photodegradation of CAP was performed in a solar simulator[6], operating at light flux of 750 W/m^2, simulating solar radiation in peak summer days.

---

[2]purchased from dr. Ehrenstorfer, 99.5% purity
[3]HP 1100 HPLC-DAD
[4]4.6 mm x 250 mm, 3 μm, Phenomenex
[5]Bruker Avance III 500 NMR spectrometer
[6]Suntest CPS+, Atlas MTT, Illinois, US

Two separate experiments were carried out. For the first one, CAP was dissolved in acetonitrile and deionized water (15 mg/L) in the volume ratio 1 : 4. In the second experiment, deionized water was replaced with tap water (11 mg/L). Solutions were irradiated in the borosilicate glass vessels together with their control samples, where the vessels were wrapped in the Al foil. The irradiation lasted continuously for seven days and aliquots were collected daily for HPLC analysis. In the collected samples four main transformation products, denoted in the further text as compounds A, B, C and H, were detected.

2.2. Preparation of degradation products

To be able to characterize the CAP degradation products, these needed to be produced in larger quantities. Low pressure Hg lamps were employed for the formation of compound A, B and C. The procedures for preparation were as follows:

- Compound A: 100 mg of CAP dissolved in 500 ml water : acetonitrile solution (4 : 1) was irradiated with four low pressure Hg lamps[7] in a quartz cell. After the total disappearance of CAP (25 min) the solvent were evaporated under reduced pressure and the reaction mixture was chromatographed by radial preparative chromatography (SiO_2, ethyl acetate - petroleum ether - acetic acid (1 : 1 : 0.05)). After crystallization (dichloromethane, heptane), 58 mg (63 %) of compound A was obtained.

- Compound B: After UVC irradiation of CAP (100 mg, 500 mg/L) to form compound A, solvents were evaporated and solid residue was dissolved in 100 mL of acetonitrile and 100 mL of 0.05 M pH 8 phosphate buffer was added. The reaction mixture was left overnight and the solvent evaporated. Solid residue was dispersed in water and filtered. After crystallization, 55 mg (59 %) of pure compound B was obtained.

- Compound C: 10 mg of compound B, dissolved in acetonitrile water mixture (1 : 1), was irradiated with 6 UVA lamps. Degradation process was monitored by HPLC and irradiation was stopped when the whole compound B was degraded to its degradation product C (10 hours). After crystallization (methanol, water), 1.5 mg (12%) of compound C was obtained. Due to its very low yield, the complete characterization of compound C could not be obtained.

- Compound H: 100 mg of CAP was dissolved in acetonitrile and deionized water (1 : 4), with added Na_2CO_3 (250 mg). The solution was

---

[7]Philips UV-C, 15 W

placed into the refrigerator for several days, until the compound H dropped out. After filtration, 67 mg (70 %) of pure compound H was obtained.

The properties of compounds A, B, C and H are listed in Table 1 and their structural formulas presented in Figure 2.

3. Results and discussion

For the study of photochemical and chemical properties of CAP and its transformation products, a mixture of water and acetonitrile was employed, due to very low water solubility of CAP and other compounds. Upon photolysis of CAP with simulated solar light (λ_{min} = 300 nm) in deionized water, the formation of one principal product (A), accompanied with two others was observed (Fig. 1, left). In addition, we investigated photodegradation of CAP with UV light in the UVC part of the spectrum (λ_{max} = 254 nm). In this case we observed nearly exclusive formation of compound A. Its fast formation was found to be due to the absorption spectrum of CAP, which peaks at λ_{max} = 200 nm (Fig. 2).
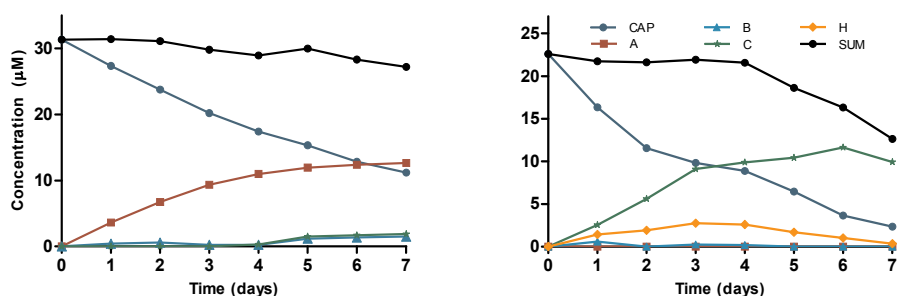

Figure 1: Degradation diagram of CAP in deionized (left) and tap water (right), irradiated with simulated solar light at high intensity. The irradiation led to formation of the three transformation products: A, B and C, however the course of the reaction is different in deionized and tap water. In tap water, additional product (compound H) was observed.

To resemble the degradation processes in natural waters, the photolysis experiment with artificial sunlight was repeated also in tap water/MeCN. In this case, quite different picture was obtained (Fig. 1, right). The compound A disappeared, the formation of formerly trace product C substantially increased and a new compound H appeared. Apparently, the buffers in tap water greatly influence the reaction pathway. A possible explanation is that compound A, as a primary photochemical product, reacts rapidly with bases in tap water, forming one of the subsequent compounds. Therefore, a solution of A in MeCN was diluted (1 : 4) with different buffer solutions; pH 4.0, 7.0 and 9.0. This test confirmed our assumption that
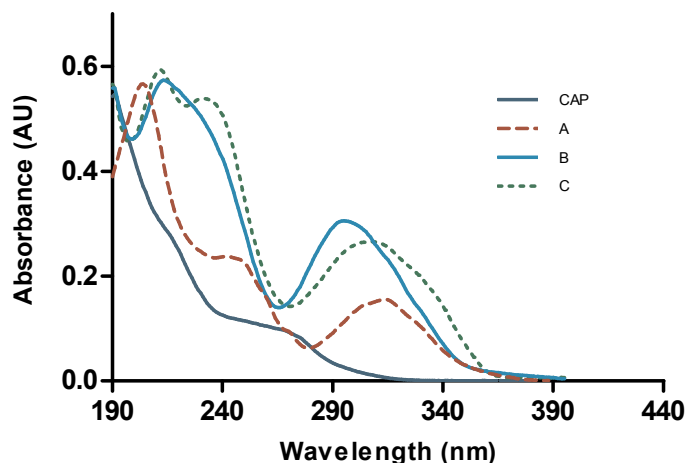
Figure 2: Absorption spectra of CAP and its photodegradation products A, B and C.

transformation of compound A to compound B is spontaneous and highly influenced by the acidity of the solution. In basic media this process was found to be rapid, in pH 7 solution it is drastically slower, while in acidic conditions the transformation was clearly inhibited and even after 6 days we were not able to detect any presence of compound B. This also explains the different course of the reaction in deionized and tap water. In deionized water, compound A remains stable due to the acidity of the media as the pH of the solution dropped from 6.1 to 3.8 at the end of the experiment. In tap water, however, compound A rapidly transforms into compound B, since the media is buffered, with a pH being around 8.2 throughout the experiment. It is also important to note that B is stable in neutral or basic aqueous solution. Despite the stability of B in neutral or slightly basic media, it was not formed in appreciable concentration during the irradiation in tap water. The formation of C under irradiation experiments is most obviously photochemical. When a solution of B in H_2O/MeCN was irradiated by UVA or UVC light, C appeared as a main product, together with several other compounds.

In control samples (wrapped in an Al foil) the concentration of CAP in deionized water remained stable throughout the experiment, while in tap water it was dropping slowly, forming the degradation product H. It can be concluded that H is formed from CAP by base catalysed reaction. When CAP was dissolved in a mixture of MeCN and aqueous Na_2CO_3, a complete conversion of CAP to compound H was observed in few hours.

3.1. Characterization of degradation products

All the above mentioned products were isolated and characterized on the basis of their ^{1}H , ^{13}C NMR and mass spectra (Table 1). During the irradiation of CAP in deionized water/MeCN, leading to A, the pH of the reaction mixture dropped substantially. A HRMS mass measurement of A showed that its molecular mass is 36 atomic units lower than the mass of CAP (Fig. 3), indicating an elimination of a HCl molecule. In the figure, various atoms are numbered for easier description of the degradation process. In ^{1}H NMR a disappearance of NH proton of the type Ar-NH-CO was observed, however, the methyl peak of a CO-NH-CH$_3$ remained as a doublet. Pyridine protons, particularly H4 in Fig 3., exhibit a moderate upfield shift, which could be caused by substitution of chlorine atom on pyridine ring by a less electron-attractive atom or group. Photochemical nucleophilic substitution of halogen in aryl halides is a well documented reaction [15] - [17] leading in aqueous solution mostly to the corresponding hydroxy derivatives. As evident from spectroscopic data, in this case the chlorine atom was not replaced by hydroxy group, therefore the most likely reaction was the intramolecular substitution by carbonyl oxygen - O9. This process seems very probable since the oxygen atom O9 is in a suitable position to form a six-membered oxazine ring. In laser flash photolysis (266 nm) on ns time scale, we have observed an instantaneous formation of a stable product, exhibiting an UV spectrum, identical to the spectrum of compound A.
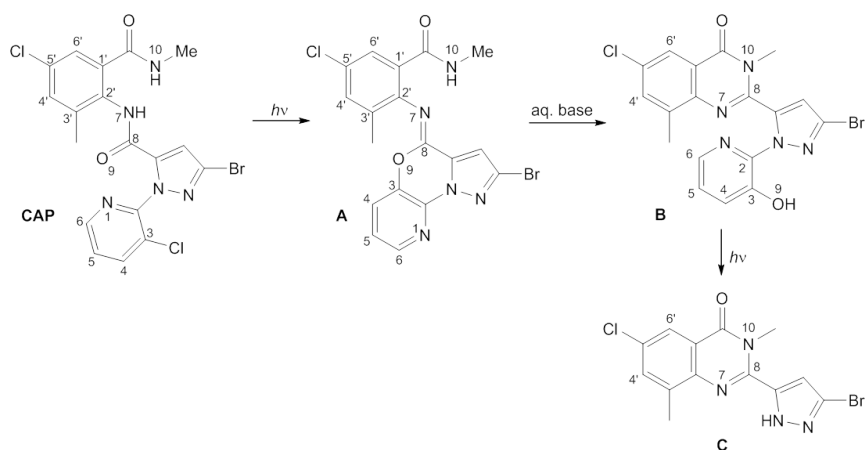

Figure 3: Photodegradaton pathway of CAP. Compound A is the first degradation pr0-oduct. Its transformation to compound B is spontaneous and is pH dependent; in basic pH it is catalysed and inhibited in acidic pH values. Under UVA, B degrades further into compound C. Various atoms in the compounds are numbered for the purpose of detailed description in the manuscript.

The formation of B from A is catalyzed by bases, present in aqueous solution; this is not a photochemical process. The HRMS measurement for

B exhibits the same molecular mass as for A, they are therefore isomers, however with very different NMR spectra. In compound A, as well as in original CAP, one of methyls is attached to NH group, resulting in a splitting the methyl signal to a doublet. In spectrum of B, the corresponding methyl exhibits a singlet, which means that the proton on the adjacent N atom is absent. The signals of protons on phenyl ring are shifted to higher δ, while those on pyridine ring to lower δ. This can be rationalized by forming of a bond C8-N10 and formation of a new quinazolinone ring with mutual scission of a former oxazine. The cleavage of the oxazine ring results in a formation of a hydroxy group on pyridine - C3. The corresponding broad singlet can indeed be observed in ^{1}H NMR spectrum of B. A further photochemical reaction of compound B leads to the cleavage of C_2-N (pyrazole) bond and expulsion of a pyridine moiety. HRMS measurement yields a molecular mass of 353. In ^{1}H NMR spectrum of compound C, the pyridine protons vanish and a new broad singlet at $\delta = 13$ appears.

4. Summary and conclusions

Since chlorantraniliprole is becoming widely used for crop protection, its monitoring in the environment is needed in order to confirm its "safe" reputation. This is valid for the the pesticide itself as well for its degradation products. In this paper we presented first results in the photodegradation pathway of chlorantraniliprole and its main direct transformation products, that have been identified and characterized. Chlorantraniliprole under simulated sunlight degrades slowly. The ratio between formed degradation products is pH dependent. Results suggest that in acidic or unbuffered neutral water media, only the first photodegradation product prevails. In basic and carbonated natural waters, the ratio of transformation is in favour of secondary photodegradation products, as well as of hydrolysis pathway of chlorantraniliprole. In order to understand the ecological risk of the pesticide, further investigations are needed to assess the toxicity and ecotoxicity of these compounds.

Table 1: Characterization of CAP's transformation products A, B, C and H.

Compound A	
Melting point	133.8 °C - 135.7 °C
^{1}H NMR (CDCl$_3$)	δ/ppm: 2.18 (s, 3H); 2.96 (d, J = 5.0 Hz, 3H); 6.17 (br s, 1H); 7.09 (s, 1H); 7.23 (d, J = 2.3 Hz, 1H); 7.26 (d, J = 2.3 Hz, 1H); 7.38 (dd, J = 4.7; 8.0 Hz, 1H); 7.85 (dd, J = 1.6; 8.0 Hz, 1H); 8.46 (dd, J = 1.6; 4.7 Hz, 1H); 10.06 (s, 1H)
^{13}C NMR (CDCl$_3$)	δ/ppm: 18.2 (CH$_3$); 26.8 (CH$_3$); 111.4 (CH); 123.5 (CH); 125.4 (CH); 127.5 (CH); 127.9 (C); 130.1 (C); 131.1 (C); 132.3 (C); 132.5 (CH); 132.8 (C); 134.8 (C); 138.4 (C); 139.6 (C); 140.0 (C); 144.9 (CH); 166.5 (C)
HRMS: M++1	446.0013; calculated for C$_{18}$H$_{14}$BrClN$_5$O$_2$: 446.0019

Compound B	
Melting point	199 °C - 200 °C
Elemental analysis	calculated for C$_{18}$H$_{14}$BrClN$_5$O$_2$: C 48.43, H 3.16, N 15.70, found: C 48.01%, H 2.72%, N 15.36%
^{1}H NMR (CD$_2$Cl$_2$)	δ/ppm: 2.48 (s, 3H); 3.33 (s, 3H); 6.75 (s, 1H); 7.07 (dd, J = 4.6; 8.2 Hz, 1H); 7.42 (dd, J = 1.5; 8.2 Hz, 1H); 7.52 (dd, J = 1.5; 4.6 Hz, 1H); 7.55 (dd, J = 0.9; 2.5 Hz, 1H); 8.18 (dd, J = 0.6; 2.5 Hz, 1H); 10.38 (s, 1H)
^{13}C NMR (CD$_2$Cl$_2$)	δ/ppm: 17.3 (CH$_3$); 32.7 (CH$_3$); 112.6 (CH); 122.7 (C); 123.9 (CH); 124.4 (CH); 127.0 (C); 127.5 (CH); 132.9 (C); 135.2 (CH); 136.6 (C); 138.3 (C); 138.6 (CH); 139.2 (C); 145.0 (C); 145.1 (C); 146.8 (C); 161.4 (C)
HRMS: M++1	446.0030; calculated for C$_{18}$H$_{14}$BrClN$_5$O$_2$: 446.0019

Compound C	
^{1}H NMR (acetone d$_6$)	δ/ppm: 2.63 (s, 3H); 3.85 (s, 3H); 7.14 (s, 1H); 7.68 (dd, J = 0.9; 2.5 Hz, 1H); 8.00 (dd, J = 0.5; 2.5 Hz, 1H); 13.22 (s, 1H)
HRMS: M++1	352.9793; calculated for C$_{13}$H$_{10}$BrClN$_4$O: 351.9727

Compound H	
Melting point	215 °C - 216 °C
Elemental analysis	calculated for C$_{18}$H$_{14}$BrClN$_5$O$_2$ x H$_2$O: C 44.58, H 2.89, N 14.45, found: C 44.71%, H 2.87%, N 14.30%
^{1}H NMR (CD$_2$Cl$_2$)	δ/ppm: 2.02 (s, 3H); 3.71 (s, 3H); 6.86 (s, 1H); 7.32 (dd, J = 4.7; 8.0 Hz, 1H); 7.43 (d, J = 1.6, 1H); 7.87 (dd, J = 1.5; 8.1 Hz, 1H); 8.01 (d, J = 2.5 Hz, 1H); 8.31 (dd, J = 1.5; 4.7 Hz, 1H)
^{13}C NMR (CD$_2$Cl$_2$)	δ/ppm: 16.7 (CH$_3$); 33.9 (CH$_3$); 112.3 (CH); 122.0 (C); 123.8 (CH); 125.8 (CH); 128.2 (C); 128.3 (C); 133.2 (C); 135.2 (CH); 139.0 (C); 139.2 (C); 140.0 (CH); 144.0 (C); 144.8 (C); 147.4 (CH); 148.7 (C); 161.6 (C)
HRMS: M++1	463.9674; calculated for C$_{18}$H$_{12}$BrCl$_2$N$_5$O: 443.9603

Bibliography

[1] Belfroid, A.C.; van Drunen, M.; Beek, M.A.; Schrap, S.M.; van Gestel, C.A.M.; van Hattum, B. Relative risks of transformation products of pesticides for aquatic ecosystems. *Sci. Tot. Environ*. 1998, 222, 167 - 183.

[2] Aga, D.S.; Thurman, E.M. Formation and Transport of the Sulfonic Acid Metabolites of Alachlor and Metachlor in Soil. *Env. Sci. Technol*. 2001, 35, 2455-2460.

[3] Martinez Vidal, J.L.; Plaza-Bolaňos, P.; Romero Gonzalez, R.; Garrido Frenich, A. Determination of pesticide transformation products: A review of extraction and detection methods. *J. Chromatogr. A*. 2009, 1216, 6767-6788.

[4] Lahm, G.P.; Cordova, D.; Barry, J.D. New and selective ryanodine receptor activators for insect control. *Bioorgan. Med. Chem*. 2009, 17, 4127-4133.

[5] Cordova, D.; Benner, E.A.; Sacher, M.D.; Rauh, J.J.; Sopa, J.S.; Lahm, G.P.; Selby, T.P.; Stevenson, T.M.; Flexner L.; Gutteridge, S.; Rhoades, D.F.; Wu, L.; Smith, R.M.; Tao, Y. Anthranilic diamides: a new class of insecticides with a novel mode of action, ryanodine receptor activation. *Pestic. Biochem. Physiol*. 2006, 84, 196,214.

[6] Cordova, D.; Benner, E.A.; Sacher, M.D.; Rauh, J.J.; Sopa, J.S.; Lahm, G.P.; Selby, T.P.; Stevenson, T.M.; Flexner, L.; Caspar, T.; et al. Elucidation of the mode of action of Rynaxypyr®, a selective ryanodine receptor activator; Ohkawa H., Miyagawa H., Lee P.W.; Pesticide Chemistry: Crop Protection, Public Health, Environmental Safety; Wiley-VCH; Weinheim, 2007, 121-125.

[7] Hanning, G.T.; Ziegler, M.; Marcon, P.G. Feeding cessation effects on chlorantraniliprole, a new anthranilic diamide insecticide, a comparison with several insecticides in distinct chemical classes and mode-of-action groups. *Pest. Manag. Sci*. 2009, 65, 969-974.

[8] Sattelle, D.B.; Cordova, D.; Cheek, T.R. Insect ryanodine receptors: molecular targets for novel pest control chemicals. *Invert. Neurosci*. 2008, 8, 107-119.

[9] Cao, G.; Lu, Q.; Zhang, L.; Guo, F.; Liang, G.; Wu, K.; Wyckhuys, K.A.G.; Guo, Y. Toxicity of chlorantraniliprole to Cry1Ac-susceptible and resistant strains of *Helicoverpa armigera*. *Pestic. Biochem. Physiol*. 2010, 89, 99-103.

[10] Lai, T.; Su, J. Effects of chlorantraniliprole on developement and reproduction of beet armyworm, *Spodoptera exigua* (Hübner). *J. Pest. Sci.* 2011, 84, 381-386.

[11] EPA. Chlorantraniliprole Chemical Documents. Pesticide fact sheet. Chlorantraniliprole. 2008.
http://www.epa.gov/opprd001/factsheets/chloran.pdf
(Access date: May 25, 2012)

[12] FAO. Pesticide report. Chlorantraniliprole. 2008.
http://www.fao.org/fileadmin/templates/agphome/documents/-
Pests_Pesticides/JMPR/Evaluation08/Chlorantraniliprole.pdf
(Access date: May 25, 2012)

[13] Xu, P.; Ren, Y.; Zhou, Z.; Liu, A.; Zhang, H. Determination of Chlorantraniliprole in Vegetables, Fruits and Grains by SPE Clean-Up and LC-UV. *Chromatographia*. 2010, 72, 763-766.

[14] Caboni, P.; Sarais, G.; Angioni, A.; Vargiu, S.; Pagnozzi, D.; Cabras, P.; Casida, J.E. Liquid Chromatography -Tandem Mass Spectrometric Ion-Switching Determination of Chlorantraniliprole and Flubendiamide in Fruits and Vegetables. *J. Agric. Food Chem*. 2008, 56, 7696-7699.

[15] Turro, N.J. Modern Molecular Photochemistry; Benjamin/Cummings: Menlo Park, USA, 1978.

[16] Klán, P.; Wirz, J. Photochemistry of Organic Compounds; J. Wiley and Sons: Chichester, U.K., 2009.

[17] Chen, C.; Yang, S.; Guo, Y.; Sun, C.; Gu, C.; Xu, B. Photolytic destruction of endocrine disruptor atrazine in aqueous solution under UV irradiation: Products and pathways. *J. Hazard. Mater*. 2009, 172, 675-684.